

# Peer Review Comments

**Article:** Eekhof, L. S., et al. (2018). Readers' Insensitivity to Tense Revealed: No Differences in Mental Simulation During Reading of Present and Past Tense Stories. *Collabra: Psychology*, 4(1): 16. DOI: <https://doi.org/10.1525/collabra.121>

**Article type:** Original Research Report

**Editor:** Monica Gonzalez-Marquez

**Article submitted:** 11 November 2017

**Editor decision:** Accept Submission

**Revision submitted:** 06 March 2018

**Article accepted:** 19 April 2018

**Article published:** 30 May 2018

---

## Responses for Version 1

### Reviewer A:

1) General comments and summary of recommendation

Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:

- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?:

I think this is a well-written, clearly structured paper that presents a study that is executed very well, in terms of design and analysis. I have some minor points that I would like to address to the authors and some suggestions and questions. However, overall I would recommend to accept this paper with minor revisions.

Important points:

- In the introduction the authors mention two times (on page 3 and 6) and later on again in the general discussion that (paraphrasing) "the long-held stance from literary studies is that the present tense is more vivid and elicits more mental simulation". However, I am not so sure about this claim. I don't think traditional literary scholars would claim this at all, seeing as how they would not isolate a textual feature like this. More likely, and more attuned to literary studies, is that there are an abundance of textual features in a literary text that when working together could be described as more or less vivid. Whether this also leads to higher mental simulation in readers is another matter, one that is much more interesting to an empirical researcher than to a traditional literary scholar. If you want to retain this claim, I think it warrants including references to relevant literature (from literary studies) that states this relationship between the use of present tense and vividness/increased mental simulation over a number of years.

- The point above ties in with my next point, namely that the fact that the “present is more vivid”-hypothesis was not confirmed based on this study, is perhaps not surprising, seeing as how there are so many other text features at play in these texts that could have influenced mental simulation or could have canceled the effect out and there did not seem to be sufficient control over these features (e.g., the two present tense stories were both written in the first person, whereas the past tense stories were written in third and first person). Obviously, this is something that the authors realize as well, but have perhaps not given enough attention to in the general discussion. I speak from experience when I say that trying to manipulate literary texts to isolate a specific effect is nearly impossible and never leads to such clear cut results as with studies that use experimenter written short texts, such as the studies the authors refer to in the discussion. This shouldn’t stop us from conducting empirical studies into the effects of literary texts, but it is something that we should be aware of and try to control for as much as possible.

- Another related point is the influence of individual differences. The authors already mention at the very start of the introduction that there are considerable individual differences in the degree to which people experience mental simulation during reading. Yet, they fail to take those individual differences into account. They did ask about reading habits and experience (frequency?), but did not take into account capacity for empathy or mental simulation in general. I would suggest coming back to this in the general discussion: could including these into the design in future studies make a difference in terms of the results? Also, another confounding variable that I was thinking of, that could have an effect in the case of their design, could be readers’ familiarity with the use of the first person perspective in the present tense. For avid readers this is probably not that much of a problem, however third person perspective and the use of the past tense are much more common in most literary genres (especially popular genres). Could this have had an influence?

- The term sensorimotor simulation refers on page 6 to “enactment of perceptual or motor experiences”, whereas on page 2 its definition also includes simulation of scenery. Was the simulation of scenery also included in your study: in the instructions to the raters and thus in the simulation score? If so, there might be a way to finetune your results by distinguishing between different kinds of simulation scores: perceptual experience, motor experience and scenery. In any case, it should be made more clear what definition of sensorimotor simulation was used in your pretest.

- I just want to make clear that the SWAS is not meant as a measure of mental simulation, but rather as a measure of narrative absorption, in which mental simulation plays a role. Therefore, the mental imagery subscale can be seen as a measure of mental simulation and perhaps the emotional engagement scale as well (I understand that the authors grasped the importance of these two dimensions over the other two, by adding items only to these two dimensions), however the transportation dimension and particularly the attention dimension are not measuring mental simulation. I think it is important to clarify this and to treat the dimensions separately in the results. In addition, it is important to use the correct terminology: now the total score of the SWAS is referred to as the mental simulation score and that is not a correct interpretation of the SWAS.

- A point related to the point above is that in terms of terminology it gets a bit confusing in the results section, especially under the heading “eye-tracking data” as the authors use the term simulation to refer both to the simulation score of the words in the text and to refer to the total score on the SWAS. An easy solution therefore would be to change the term given to the total SWAS score to distinguish it from the text score and to be mindful of the SWAS’s origin.

- The last point: The rationale that mental simulation takes time and effort applies to a lot of processes involved in fiction reading. How can you be sure that slowing down reading is purely connected to mental simulation processes? As I suggest in another point, could it also be comprehension processes at play? Perhaps you could expand on this rationale a little more with some relevant literature and possible alternative explanations of slowed down reading.

Clarifications needed:

- The last sentence on page 10 seems to imply that there was a fixed order to the way in which the tenses were presented to the readers, even though the stories were randomized.
- In your method section it says that in the pretest the raters underlined words. However, I am wondering what the exact instructions were to the raters. Were they only allowed to underline isolated words or were they allowed to underline phrases and sentences and did you then score the words individually again? I think this is important, because I can imagine that a lot of sensorimotor imagery descriptions happen on the sentence level rather than the word level. If the raters underlined sentences rather than words, how did you treat underlined words that in essence have nothing to do with mental simulation, but are just part of a sentence that does influence mental simulation. Are all words with a high simulation score equal? And if not, are there ways in which you can take them into account in your analyses in ways that differentiate them. Perhaps by finetuning your simulation rating system, you could finetune your results.

Suggestions/questions:

- You did include some comprehension questions into your questionnaire. If there was significant variance between participants on these questions you could perhaps check for a possible alternative explanation for the result that participants read the past tense story slower. So in other words: try to match the degree to which subjects slowed down in past tense stories to the offline measure of comprehension.
- I would be curious to see what happens if you run an exploratory factor analysis on the SWAS plus the additional items on the imagery and emotion dimensions. Perhaps you would find different forms of simulation like you suggest: mentalizing and sensorimotor simulation.
- Could you check the relation between the SWAS scores and the simulation scores for the text in any way? In any case you could check the relationship between the SWAS scores and the eye tracking data per dimension on the SWAS while taking into account comprehension. These are, by the way, just suggestions to further dig into your data and perhaps finetune your results, I do not want to suggest that they are necessary revisions.

Minor points:

- When introducing the example sentences on page 4, I think you could be clearer by stating "For example, such theories would suggest that the small discourse in sentence 2 below..."
- I think it is the "Author Recognition Test" not "Author Recognition Task".
- It is not clear what the "adjective likert scales" are referring to in the results section: they were not introduced in the method section. Also, I believe with a Cronbach's alpha of .76 you can speak of "reliable enough" (the border lies more at the .65 level).

2) Figures/tables/data availability:

Please comment on the author's use of tables, charts, figures, if relevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here):

Table 1: I think a better translation for the Dutch word "De Invaller" would be "The substitute"

Table 2: I think the fact that "amount of words changed" is added to this table is a little confusing,

as this information pertains to the manipulated versions of the stories and not the original stories (that the rest of the table pertains to). In relation to this a question: did the amount of words changed in the stories influence the average amount of words underlined in the manipulated versions of the stories?

### 3) Ethical approval:

If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.

If not, please detail where you think a further ethics approval/statement/follow-up is required.:

Ethical approval is adequately addressed.

### 4) Language:

Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:

I think in general the English in this paper is very good. Some small issues:

- on page 11 there is one sentence in the last paragraph of the procedure section printed twice
- on page 12 the comma between "potential" and "future" should be removed
- i.e. and e.g. should consistently have comma's behind their periods, like so: i.e., or e.g.,
- on page 20 it should read "The present is more vivid" hypothesis

-----

### **Reviewer B:**

#### 1) General comments and summary of recommendation

Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:

- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?:

Overall, the manuscript is clear and concise. The research methods are for the most part sound, despite a few clarifications and inclusions, which I believe are necessary for publication. The focus of

my review is on four problems that I see with the manuscript in its current form; the equivocation of mental imagery versus mental simulation, sample size, effect size reporting, and model fitting.

First, I would like the authors to include a few sentences on the difference between mental imagery and mental simulation in the introduction. According to their hypothesis, mental simulation requires time and effort, "The rationale here was that mental simulation takes time and effort, and should therefore show up in the eye-tracking data as increased reading times for passages that elicit simulation." (p6). My concern here is on the use of the word "effort". Are they talking about conscious effort? If so, that would seem to be more along the lines of mental imagery rather than simulation. Perhaps the authors could clarify.

Second, according to the project wiki on the OSF, between 06-04-2017 and 20-04-2017 a change was made to the main analysis section to include a Bayesian Sequential Analysis. However, there is no such analysis in the manuscript, nor is there a power analysis of any kind. I am not sure how the authors decided on the number of participants for their study. To estimate the effect size for participants, I used Carreras, Carriedo, Alonso, and Fernández (1997), a well-known study on tense effects in discourse, and cited by Eerland et al. Using an equation from Lakens (2013), I found the F1 ANOVA statistic ( $F(1, 28) = 11.58$ ) for the main effect of tense in experiment 1 (E1) of Carreiras et al., calculated a partial eta-squared  $(11.58 \cdot 1) / (11.58 \cdot 28 + 1)$ , and found an effect size of  $\eta^2 = 0.03$ . For experiment 2 (E2), where they reported  $F(1, 45) = 6.93$ , I calculated the effect size again  $(6.93 \cdot 1) / (6.93 \cdot 45 + 1)$ , and found that it was  $\eta^2 = 0.02$ . Both effect sizes were very small, but significant at an alpha level of 0.05. Given these small effect sizes, I worry if the sample size chosen by the Eekhof et al. was adequate. Would the authors please discuss how they decided on their sample size?

Also, with regards to sample size, I want to discuss the number of items, or texts, in Eerland et al. Again, I am going to draw on the effect sizes that I calculated from the item analyses (F2) for the main effect of tense from Carreiras et al. (1997). In E1 the effect size was  $\eta^2 = 0.02$ . This effect was not significant; reported as  $p > 0.05$ . In E2 it was  $\eta^2 = 0.02$  and significant ( $p < 0.05$ ). The number of items, or short paragraphs, used in E1 and E2 of Carreiras et al. was 51 and 67 respectively. However, the number of items in Eekhof et al. was only 4, which contained from 743 to 2016 words. It is not clear how many verbs were in each story, and that seems to be important. It would be good, in my opinion, for the authors to clarify the number of verbs, and discuss how they decided on the number of items.

Third, I would like the authors to report an effect size. They wrote, "We do not report effects sizes given the lack of consensus on how to report these for linear mixed models." (p13). However, I would say that unless there is a published article that I am unaware of where different alternatives for reporting effect sizes for linear mixed-effect models (LMMs) exists, then this sentence should be reworded. If there is one, then the authors can simply cite it here to support the argument that there is a lack of consensus. However, I think the problem is not a lack of consensus, but instead a lack of any published article on how to report effect sizes for LMMs. One of my personal views is that if an article showing null effects is to be published, then it should report an estimate of the effect size to aid in future research and meta-analyses. My recommendation is that the authors calculate and report  $R^2$  as discussed in the only article that I know of that discusses the problem of effects size estimates for GLMMs (Nakagawa & Schielzeth, 2013). Alternatively, the authors could report F1 and F2 ANOVAs with partial eta squares. Given the small number of items, I do not see that LMMs are particularly helpful anyway.

"To test for statistical significance of effects, Likelihood Ratio Tests were used comparing the full model with the model without the effect of interest." (p13)

I would recommend clarifying that the "full model" included only random intercepts, but not random slopes. It can be confusing because some analysts might call them "empty models", and to those peo-

ple “full models” include both random slopes and intercepts. Maybe just say “comparing a model with all fixed effects and both random intercepts to…”

Carreiras, M., Carriedo, N., Alonso, M. A., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition*, 25(4), 438-446.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 1-12.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142.

## 2) Figures/tables/data availability:

Please comment on the author’s use of tables, charts, figures, if relevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here):

The tables and figures are relevant. The data sharing has already been improved in this round of reviews, and is adequate.

## 3) Ethical approval:

If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.

If not, please detail where you think a further ethics approval/statement/follow-up is required.:

Ethical approval was stated. I do not believe that this is a concern.

## 4) Language:

Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process.:

My opinion is that the flow of the arguments in the manuscript is clear and jargon free.

-----

## **Reviewer C:**

### 1) General comments and summary of recommendation

Describe your overall impressions and your recommendation, including changes or revisions. Please note that you should pay attention to scientific, methodological, and ethical soundness only, not novelty, topicality, or scope. A checklist of things to you may want to consider is below:

- Are the methodologies used appropriate?
- Are any methodological weaknesses addressed?
- Is all statistical analysis sound?
- Does the conclusion (if present) reflect the argument, is it supported by data/facts?
- Is the article logically structured, succinct, and does the argument flow coherently?
- Are the references adequate and appropriate?:

I think this manuscript undoubtedly has potential, but there are some issues which I would like to see addressed before I can recommend accepting the article for publication. The work is methodologically sound; there were clear hypotheses, which were preregistered, and the data are in principle able to address those hypotheses. However, I have some concerns about the data analysis.

First, I would like to know why no random slopes were included for Simulation, Tense and Simulation  $\times$  Tense per story and Tense and Simulation  $\times$  Tense per subject. Based on some additional models I've fitted it seems that the data allow for a more comprehensive random effects structure than the one used in the final model, and including random slopes might lead to different conclusions about the main effects of tense and simulation. Bates et al. (2015; <https://arxiv.org/abs/1506.04967v1>) provide some useful guidelines on how to select the most comprehensive random effects structure that is justified by the data.

Second, it is not clear to me why the authors decided not to perform Bayesian data analysis like planned in the preregistration. Arguably LMEs are indeed more suitable and precise for eye-tracking data (as is claimed in the Data Accessibility Statement) than Bayesian ANOVAs, but there are also Bayesian alternatives to the frequentist linear mixed models, which are no less suitable for eye-tracking data. It should be relatively straightforward to transform the LME model into a Bayesian model using, for example, the `stan_lmer()` function from the `stanarm` package in R. A Bayesian analysis would be better suited to support the conclusions that the authors make about the lack of an interaction effect, since strictly speaking absence of evidence is not evidence of absence in the frequentist framework. In contrast, a Bayesian analysis would allow the authors to draw conclusions about how likely it is that the null hypothesis is true, given their data.

Third, why were the words only scored for sensorimotor simulation in the original version and not in the manipulated version? It was expected that there would be differences in how vivid the language would be perceived, depending on tense, so it would be useful to have scores for both versions of each story. And how was the judges' task formulated?

Finally, some minor comments:

p. 10 What would count as "below chance"? Fewer than 2 correct responses?

p. 11 "As a final question..." repeats the previous sentence.

p. 11 How did the excluded participant's reading behavior diverge from normal reading?

p. 13 From which corpus were the lemma frequencies taken? Are they genre-specific, or at least

modality specific?

P. 13 Another predictor, apart from frequency and word length, which seems to explain a non-negligible amount of variance is page number: participants speed up quite a lot across the course of the story.

2) Figures/tables/data availability:

Please comment on the author's use of tables, charts, figures, if relevant. Please acknowledge that adequate underlying data is available to ensure reproducibility (see open data policies per discipline of Collabra here):

The tables and figures are clear.

3) Ethical approval:

If humans or animals have been used as research subjects, and/or tissue or field sampling, are the necessary statements of ethical approval by a relevant authority present? Where humans have participated in research, informed consent should also be declared.

If not, please detail where you think a further ethics approval/statement/follow-up is required:

Informed consent was declared, and there are no ethical concerns.

4) Language:

Is the text well written and jargon free? Please comment on the quality of English and any need for improvement beyond the scope of this process:

The text is well written.

---

### **Editor Decision for Version 1**

**Editor:** Monica Gonzalez-Marquez

**Affiliation:** RWTH – Aachen University, DE; Cornell University, USA

**Editor decision:** Revisions Required

**Decision date:** 18 January 2018

Dear Mrs Lynn S. Eekhof,

After review, we have reached a decision regarding your submission to Collabra: Psychology, "Readers'

insensitivity to tense revealed: no differences in mental simulation during reading of present and past tense stories". Our decision is to request revisions of the manuscript prior to acceptance for publication.

The full review information is included at the bottom of this email. There may also be a copy of the manuscript file with reviewer comments available once you have accessed the submission account. A summary of the requested edits from the editorial team can be found below. Please consider these points and revise the file accordingly:

#### Editorial Revision Requests:

The reviewers were generally quite positive about this manuscript, although they did have several general concerns as briefly summarized here.

1. Reviewers A & B both had questions about simulation and mental imagery, albeit from two different perspectives, reviewer A from narratology and B from psycholinguistics.
2. Reviewer A also required further explanation of your use of the SWAS instrument.
3. Reviewers B & C both had questions about the differences in the analyses proposed in the registered report and those described in the manuscript, specifically with regard to the proposed Bayesian sequential analysis.
4. Reviewer B also asked for an explanation for not reporting effect sizes beyond that already provided in the manuscript.
5. Reviewer C inquired about your motivations for not using random slopes for Simulation, Tense and Simulation  $\times$  Tense per story and Tense and Simulation  $\times$  Tense per subject.
6. In addition, Reviewer C asked why words were scored for sensorimotor simulation in the original version but not in the manipulated version.

The reviewers also have a series of minor revisions, which can be found in the copies of their reviews below.

To access your submission account, follow the below instructions:

- 1) login to the journal webpage with username and password
- 2) click on the submission title
- 3) click 'Review' menu option
- 4) download Reviewed file and make revisions based on review feedback
- 5) upload the edited file
- 6) Click the 'notify editor' icon and email the confirmation of re-submission and any relevant comments to the journal.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing; therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please do contact us.

Please could you have the revisions submitted by 2 February 2018. If you cannot make this deadline, please let us know as early as possible.

Kind regards,

Monica Gonzalez-Marquez

RWTH – Aachen University

monica.gonzalez@ifaar.rwth-aachen.de

---

### Author's Response to Review Comments for Version 1

**Author:** Lynn S. Eekhof

**Affiliation:** Centre for Language Studies, Faculty of Arts, Radboud University Nijmegen, NL

**Revision submitted:** 06 March 2018

Dear dr. Gonzalez-Marquez,

Thank you for handling our manuscript Readers' insensitivity to tense revealed: No differences in mental simulation during reading of present and past tense stories. We greatly appreciate the reviewers' feedback and have done our best to incorporate their comments in a revision of the manuscript. Please find our responses to the points the reviewers raised in the attachment. In line with changes made to the analysis we have also uploaded new files on the Open Science Framework (<https://osf.io/qynhu/>).

We hope that the improved manuscript will be suitable for publication in Collabra Psychology's new research Nexus Verb Aspect and Event Cognition.

Kind regards,

Lynn Eekhof, Anita Eerland, and Roel Willems

**Attached document:**

[collabra-4-121-pr\\_Auth\\_Resp.docx](#)

---

### Editor Decision for Version 2

**Editor:** Monica Gonzalez-Marquez

**Affiliation:** RWTH – Aachen University, DE; Cornell University, USA

**Editor decision:** Accept Submission

**Decision date:** 19 April 2018

Dear Mrs Lynn S. Eekhof,

After review, we have reached a decision regarding your submission to *Collabra: Psychology*, "Readers' insensitivity to tense revealed: no differences in mental simulation during reading of present and past tense stories", and are happy to accept your submission for publication, pending the completion of copyediting and formatting processes.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. The accepted submission will now undergo final copyediting. You will be contacted once this is complete to answer any queries that may have arisen during copyediting and to allow a final chance to edit the files prior to typesetting. If you wish to view your submission during this time, you can log in via the journal website.

The review information should be included in this email.

Kind regards,

Monica Gonzalez-Marquez

Cornell University, USA

monica.gonzalez@ifaar.rwth-aachen.de