Reviewers' comments.
Responses in green.

Reviewer #1:
Manuscript Summary:
This manuscript presents methods that could prove useful to many people who study visual attention.

My review focuses on the Bayesian analyses and the R scripts (and some minor editorial remarks). I leave it to other reviewers to assess the content regarding the TVA model.

Major points are marked by triple asterisks, ***. Minor points are marked by single asterisks, *.

Suggestions for the interpretation of the Bayesian analyses.

Thank you for reviewing the manuscript. We improved the manuscript concerning the issues raised in your comments as described below for each point.

(1)
*** Top of p. 10: "... 0, no difference, is just at the fringe [of] the 95% HDI, hence it is highly unlikely, but there is a larger-that-5% probability that the real difference is zero." This is not correct as stated. For a continuous posterior distribution, probability statements can only be made about intervals of parameter values. Presumably what is intended is a statement about the proportion of the posterior distribution that is near zero, for some interval near zero. Thus, what is needed is the specification of a ROPE (region of practical equivalence, see Kruschke, ref. 13) around zero. Then a statement can be made about the proportion of the posterior distribution in the ROPE around zero. (If the manuscript was trying to indicate how much of the posterior distribution was below 0, then the statement should be phrased as such. But I do not think that information has any value. Saying that most of the posterior distribution falls on one side of 0 does not indicate how far from 0 it falls. A ROPE is needed to decide that the deviation from 0 is meaningful.)
Thanks for the clarification. The sentence has been changed to not claim the "larger-than-5%" part. See also response to next comment …

(2)
*** In general, when interpreting continuous posterior distributions, it's best to focus on their central tendency and HDI, because those are direct summaries of the posterior distribution. If an analyst wants to go further and make a decision about whether or not to reject a difference of zero, then a ROPE must be specified. It would be useful for the manuscript to discuss what would be a reasonable ROPE

around zero. (Or, instead the analyst can venture into Bayesian hypothesis testing with Bayes factors, but that requires further careful specification of meaningfully informed null and alternative hypotheses.)

At several locations in the text, "point estimate" has been changed to "central tendency" to highlight that the reported number describes a distribution. Furthermore, a paragraph briefly explaining the concept of ROPEs was added to the discussion, including an example how a rope could be set up for differences of TVA-based estimates.

(3)

*** P. 16, 2nd paragraph: "In the Bayesian framework suggested here, it is allowed to ... collect further data without ... increasing the chances of false alarms." This is not correct. False alarm rates for hypothesis testing are influenced by the stopping rule, even in a Bayesian framework. See e.g., Ch. 13 of Kruschke (ref. 13), and more recent articles in the literature such as Schönbrodt, Wagenmakers, Zehetleitner, and Perugini (2015) Psychological Methods. The issue of optional stopping is subtle and shouldn't be dismissed so easily.

As the issue would require further explanation (and the manuscript is already at JoVEs length limits; especially after adding the ROPE section), the claim was now simply removed. Kruschke's book and other resources are referenced throughout the article as a source for interpreting Bayesian statistics for the interested reader.

Suggestions for the R scripts: (I only tried a few of the R scripts.)

First of all, many thanks for trying out the scripts and suggesting improvements.

(4)

*** The assumptions of the power analysis need a more explanation. In particular, how is the hypothesis specified? In the script power-exp1.R, what are the mysterious variables SOAs, reps, C_mean_sd_subsd, w_probe_mean_sd_a, w_probe_mean_sd_n, and so on? The comments in the R script are not sufficient to understand what is going on.

The comments have been improved. However, highly common terms (as SOA) have not been explained within the scripts.

*** In the scripts power-exp1.R and In the script power-exp2.R, line 62 must be changed. Currently, line 62 is

print(nSimulations)

which only presents the total number of simulations to be run. Instead, it should be changed to something like this:

print(paste("Completed Simulation" , sim , "of" , nSimulations))

This was a leftover debugging statement and was removed now. The number of completed simulations is shown in the (now improved) output.

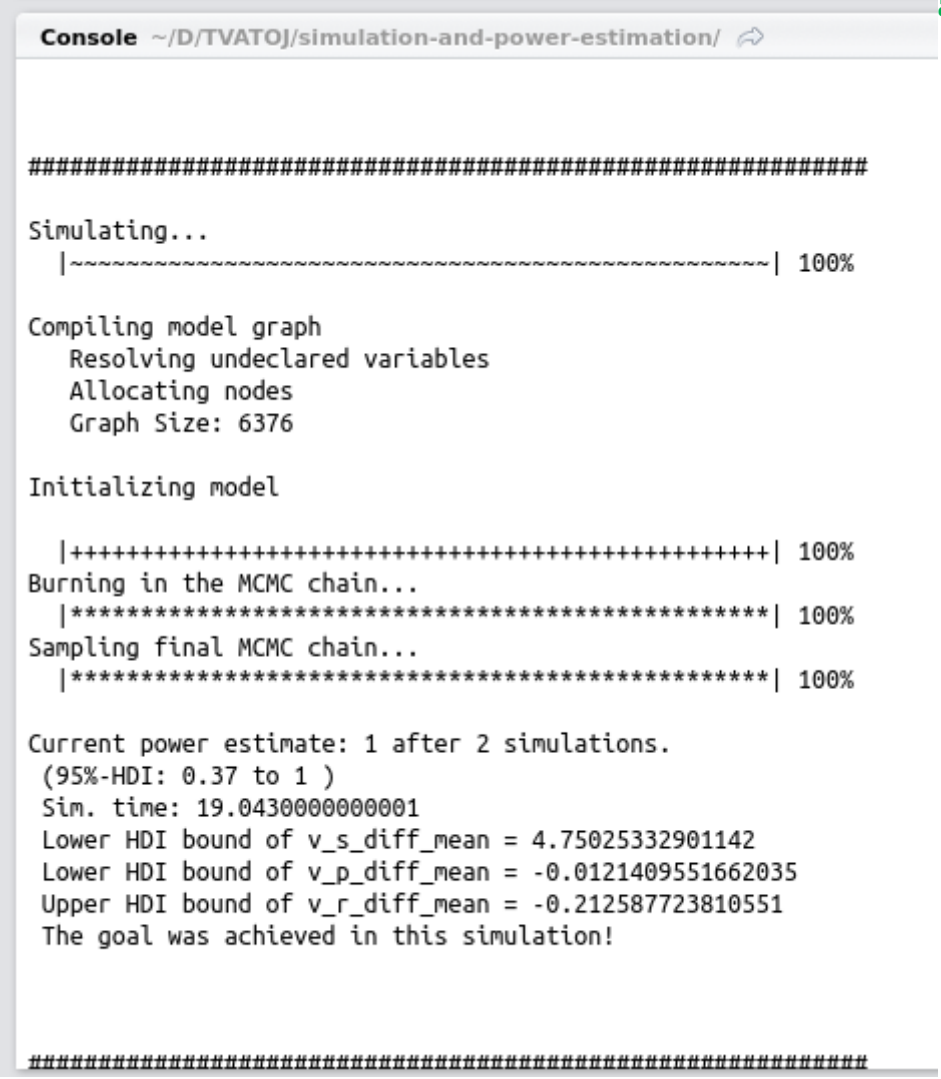
* In the scripts power-exp1.R and In the script power-exp2.R, something seems to be wrong with the production of the output line consisting of "~~~~~~~..." followed by "100%". I assume it's supposed to be representing the proportion of simulations completed, but it's not functioning properly.

I was not able to reproduce the problem. The "~~~" indeed are a progress bar which works for me in Rstudio and command line R. I have the suspicion, however, that a missing line break before the progress bar could have lead to the problem on some systems. Now the output was improved (inserting breaks above and below the progress bar). See image below.

(7)

* It would be nice if the power scripts produced an ongoing summary of the estimated power on screen, instead of putting the progress only in a saved file.

This was added:



(8)

* The R scripts need to be better formatted so that they display nicely when opened in RStudio. Presently, the lines are not formatted for standard 80-character wide

displays; consequently wrapped lines are very difficult to visually group and section indentation is not evident. If these R scripts are supposed to be used by others, they should be more reader friendly.
The code has been reformatted. We would be glad if it is helpful to others. It still does not have the best usability, but should be easier to read and understand now. We are currently working on highly usable python-based framework to replace the R scripts in the future.

(9)
* R code: It's good practice to use full TRUE and FALSE and not just T and F, because the single letters T and F can be inadvertently used as variable names and given other values.
Thanks for pointing this out. It was fixed.

Suggestions for the figures:

(10)
*** Figure 1 is redundant at best and confusing at worst. It should be omitted entirely. The diagram is meaningless without the list of equations at its side, and the content of the equations is not elucidated by the diagram. Moreover, the diagrammatic conventions are not explained (circle vs square vs double circle, gray vs white, plates, etc.). Perhaps retain the list of equations with some explanatory text to provide the details of the model, but omit the diagram.
Thank you for pointing out the difficulty with this figure. We find the diagram rather helpful; it was intended to illustrate the hierarchical structure in the model (group->subject->SOA). Therefore, we kept the diagram but extended the explanation in the figure legend to improve on this issue.

(11)
*** In Figures 2, 3, and 5, the sigmoidal curves need their axes to be labeled!
This has been corrected.

(12)
Suggestions for spelling or grammar:
Thank you for pointing out these mistakes; they have been corrected. Additional language improvements have been made.

* The first time "TOJ" is used in the main text it has not yet been used in its expanded form in the main text. That is, the expanded form only occurs in the abstract.

* Page 3, margin line 92: Insert "stimulus" --- "... processing rate of the attended STIMULUS ..."

* page 4, equation 4: Not "else". Instead, "otherwise".

* page 10, final line: Not "intendedly". Instead, "intentionally".

* page 11, line 411: "compared TO the control condition"

* page 15, line 584: Not "mathematically model". Instead, "mathematical model".

* Figure 4, horizontal axis: Not "requiered". Instead, "required".

Thank you for the opportunity to review this manuscript.

Major Concerns:
N/A

Minor Concerns:
N/A

Additional Comments to Authors:
N/A

Reviewer #2:
Manuscript Summary:
The authors describe how to design, carry out and analyze temporal order judgements where participants have to judge which of two presented stimuli came first. The demonstrate how the use of Bundesen theory of visual attention (TVA) can be used to gain more interpretory security about the underlying processes determining TOJ (i.e. processing speed for stimulus one and two).
The presented contents are scientifically highly interesting and especially the combination of TOJ experiments with the presented computational analyses are innovative. It is convincingly stated that TOJ experiments provide a critical advantage compared to standard TVA-based whole and partial report experiments in that they are not limited to the presentation of verbal stimuli. On the other hand, TOJ experiments, although influential in psychological research, have a central weekness that is well documented by the authors. That is, without the use of computational modelling provided by TVA, it can not be answered whether the attended stimulus is processed faster or the unattended stimulus is slowed. This question is central and can be solved by this critical combination.
Interested viewers will be able to replicate these experiments based on highly appropriate explanation and visualization. Examples for representative

experimental questions and related results are given. Viewers will, on the basis of these example, be able to generate a wide range of new research questions and I think that the video can inspire such studies.

Major Concerns:
I do not have any major concerns.

Minor Concerns:
The loudness of the speakers varies.

Additional Comments to Authors:
N/A

Thank you for reviewing the manuscript and pointing out the loudness issue. We will improve on this when submitting the high-quality version of the video to the journal.