PCI Registered Reports #808

One and only SNARC? Spatial-Numerical Associations are not fully flexible and depend on both relative and absolute magnitude

Stage-2 Submission, Version 2

Dear Rob,

Thank you for reviewing our Stage-2 Registered Report so quickly and for finding a new reviewer! We appreciate it and are grateful for your helpful feedback.

Please find our replies to the comments in the following. We revised the manuscript accordingly and highlighted all related changes.

During the revision process, we have noticed a mistake in our data analyses. Namely, when calculating the Bayes Factors for t-tests, we have accidentally not changed the default *independent* t-tests to *paired* t-tests in the BayesFactor R package. This means that we *underestimated* most reported Bayes Factors; i.e., most Bayes factors are even more convincing than previously reported. Nevertheless, we are sorry that this has happened, but glad that we have noticed this mistake before the paper gets published. This is the reason for many changed Bayes Factors in the revised manuscript and for many tests resulting in even stronger Bayesian evidence than before. Note, however, that no results changed substantially as regards their interpretation – if anything they got more somewhat convincing.
When analyzing the data with *paired* t-tests instead of *independent* t-tests, we noticed that the we could have stopped data collection with the minimum sample size of 200 in Experiment 2 as well (just as in Experiment 1), because the optional stopping criterion would have been reached earlier than at a sample size of 300. Thus, we inserted a footnote with the results taking into account only the first 200 out of the 300 participations. Note that the results would not have been substantially different with 200 instead of 300 participants.

We believe that your and the reviewer's comments have significantly improved our manuscript. Additionally to your and the reviewer's suggestions, we have made two more improvements: First, to meet the demand for a summary of all hypotheses and outcomes, we have created a new figure (see Figure 5 in the revised manuscript). Second, we have made Figures 3 and 4 more accessible (e.g., for color-blind readers) by using a solid line for the lower and a dashed line for the higher number ranges.

Best wishes,
Lilly (on behalf of all authors)

Congratulations on completing this fascinating experiment, and commendations on the very clear and thorough write-up of your findings. The Stage 2 review process for Registered Reports is typically quite light touch, provided that the authors have not modified the Stage 1

components (which you have not), and have followed their pre-registered analysis plan (which you have).

I have received helpful comments from one expert reviewer (Claudia Gianelli), but was not able to get both original Stage 1 reviewers to return. I have read over the Stage 2 report myself, and noted only a few very minor issues to attend to (in addition to considering Claudia Gianelli's suggestions):

1) In your footnote 2 on page 28 (within Results), I think that you may under-emphasise the importance of the differences in outcomes that follow from the exclusion of numbers 1 and 4 (from the lower and higher ranges respectively). What you have written if perfectly correct, but the reader may miss that the "… did not change substantially…" applies specifically to H1a, and may overlook the later statement that the change made the evidence become inconclusive for H2b and H3. Just a subtle change of wording here would likely avoid such a misreading, and make the reporting of Results more congruent with the discussion of this issue on p44).

Thanks, what you write makes perfect sense. We have changed the footnote accordingly and added the parts that are underlined in the following:

"When excluding number 0 from the lower and number 4 from the higher range in the analyses, the results of the confirmatory data analysis did not change substantially for Hypothesis 1a with $BF_{10} = 2.09 * 10^9$ and for Hypothesis 1b with $BF_{10} = 6.65 * 10^5$, and naturally remained the same for the dRT of number 5 in Hypothesis 2a. However, for Hypotheses 2b and 3, the exclusion of number 0 from the lower and number 4 from the higher range led to inconclusive evidence regarding AMdependency of the number mapping on the MNL and of the strength of the SNARC effect, with $BF_{10} = 1.71$, and $BF_{10} = 0.62$, respectively. This stands in contrast with moderate evidence against AMdependency found when including number 0 in the lower and number 4 in the higher range."

2) Your focus on reporting results is on statistical outcomes (BF's), but it would be very informative to provide estimates of effect size for the RM and AM components of the SNARC, both in terms of raw effect sizes (dRT in ms) and a suitably standardised measure (e.g. Cohen's d). Within the Discussion, you should pick up upon the relative strength of the RM and AM effects. This would make it clear that the RM effect is massively larger than the AM effect. I do not think this takes away at all from the importance of your finding that AM also matters, but providing a concrete sense of scale (beyond BF's) would be helpful, and should inform your discussion (and Abstract).

Thanks for this helpful suggestion. In the Method part, we inserted (page 25): "Additionally to all effects reported in the unit of interest, we provide effect sizes in terms of Cohen's d for all Bayesian t-tests. Effects of $d \geq 0.2$, $d \geq 0.5$, or $d \geq 0.8$ will be interpreted as small, medium, or large effect sizes, respectively." We have thus calculated Cohen's d for all hypothesis tests and added them to the "Results" part of the manuscript.
Also, we refer back to them in the "Discussion", saying how large each effect was. We draw the following conclusions from the effect sizes (page 46): "Note that in contrast to

the small effect size in both ranges in Experiment 1 ($d = 0.38$ and $d = 0.47$), the effect size was medium in both ranges in Experiment 2 ($d = 0.71$ and $d = 0.52$). Thus, our results support the claim that the inclusion of number 0 in the stimulus set or a potential confound with the MARC effect due to an unequal number of odd and even numbers might have decreased the SNARC effect in the seminal studies by Dehaene et al. (1993) and Fias et al. (1996)."

Moreover, we state that the (page 47) "differences indicating AMdependency are only small ($d = 0.27$ and $d = 0.28$ for Hypotheses 2b and 3), but so are the differences indicating RMdependency ($d = 0.42$ and $d = 0.24$ for Hypothesis 2a)."

3) When you report "strong evidence for a correlation" between SNARC slopes in Experiment 2, it might be helpful to specify, "strong evidence for a modest correlation", because a reader who was not paying sufficient attention might assume that you mean evidence for a strong correlation. Again, what you have already written is perfectly correct; this is just to guard against a possible misreading.

We have made the change you suggested on pages 34 and 47, writing that Experiment 2 revealed strong evidence for a _moderate_ correlation, with an estimate of $r = 0.34$.

4) When a BF has seven digits before the decimal, it feels like spurious precision to report digits beyond the decimal place. I would be tempted to round such large BF's to integers. However, I also recognise that 2 dpl is the convention you have used elsewhere and so it is fine to keep it this way if you prefer.

Thanks we agree with you. In the revised manuscript, instead of writing up each BF with up to seven digits to a precision of two decimals, we now write only each BF with up to four digits including two decimals. All BF that have at least five digits are abbreviated with a multiplication of a power of 10, which we previously did only for BF with at least eight digits.

Please note that I will be away now until August 5th, but would be pleased to deal with a final version on my return. If you submit it sooner than that, then it may be dealt with by another member of the PCI RR board (which is fine by me too – I do not think these final revisions are in any way problematic).

Best wishes,
Rob

Thanks, we are happy to hear that you consider our stage-2 registered report as a solid addition to the SNARC literature and appreciate your helpful suggestions.

We have created a Readme file that can now be found in the main OSF project (see https://osf.io/z43pm/files/osfstorage), which includes the following information:

"The main OSF project "SNARC Flexibility: Range (In)Dependency" (https://osf.io/z43pm/) contains four ZIP folders called "Exp1_SNARC_RangeDependency_DataAnalysis" (version 1) and "Exp1_SNARC_RangeDependency_DataAnalysis_v2" (version 2 with some corrections and extra analyses) for Experiment 1, and "Exp2_SNARC_RangeDependency_DataAnalysis" (version 1) and "Exp2_SNARC_RangeDependency_DataAnalysis_v2" (version 2 with some corrections and extra analyses) for Experiment 2. Each of them contains (1) a subfolder with the data (200 files for Experiment 1 and 300 files for Experiment 2, that is one file per participant), (2) an R script for the data preprocessing and analysis, and (3) a subfolder with the outputs (txt files with results, pdf files with plots).

The main OSF project also contains a folder named "Conferences, presentations, etc.". This folder includes (1) a poster presented by Lilly Roth at the MCLS conference in Antwerp, Belgium, in 2022 (before submission of the Stage 1 Registered Report), and (2) slides for a talk held by Lilly Roth at the MCLS conference in Washington D.C., USA, in 2024 (after submission of the Stage 2 Registered Report).

There are two components associated with the main OSF project. These are (1) "PCI Stage 1 Registered Report Materials" and (2) "PCI Stage 2 Registered Report Materials". Each of them contains separate folders for each version of the Registered Report that was submitted to PCI. In each version, the manuscript (Introduction and Method for Stage 1; Introduction, Method, Results, and Discussion for Stage 2) was

uploaded as a pdf file together with the supplementary material such as power considerations and a study design table.

The present study included two experiments. Experiment 1 was a direct replication of Dehaene et al.'s (1993) and Fias et al.'s (1996) studies, testing the SNARC effect in the number ranges 0 to 5 vs. 4 to 9. The demo version for Experiment 1 can be found at https://luk.uni-konstanz.de/numcog_3/?demo&e1. Experiment 2 was a conceptual replication investigating the SNARC effect in the number ranges 1 to 5 (excluding 3) vs. 4 to 8 (excluding 6). The demo version for Experiment 2 can be found at https://luk.uni-konstanz.de/numcog_3/?demo&e2."

2. Reporting of data collection and pre-processing is suboptimal at the moment. The authors list all rejections and state that some participants were rejected for multiple reasons, but it is hard to tell how many and at which stage of pre-processing.

In the revised version, we explicitly state in the "Data preprocessing" paragraph of the "Method" part. (pages 21 and 22): "The preprocessing steps were applied in the exact order as they are listed in the following."

Moreover, we have restructured the "Data preprocessing paragraphs" in the "Results" part (pages 28 and 29 for Experiment 1 and pages 32 and 33 for Experiment 2). Preprocessing steps are now listed in the exact same order as previously described in the "Method" part, which is the order they were applied in.

3. The rationale behind exploratory analyses should be made more transparent to the reader, even the one less familiar with the topic. At the moment, exploratory analyses start right the confirmatory ones quite abruptly.

Thanks for raising this point. We have made the rationale behind the exploratory analyses more clear in the revised manuscript. Please find the added parts underlined and the removed parts striked through in the following:

Revised results for exploratory analyses in Experiment 1 (page 31):

"In addition to the confirmatory data analyses and in order to disentangle the possible scenarios illustrated in Figures S1 and S2 and Table S1 (see Supplementary Material), the mean-number intercepts were compared between ranges. The mean number in the lower range from 0 to 5 was 2.5 with a dRT estimate of 2.09 ms, and the mean number in the higher range from 4 to 9 was 6.5 with a dRT estimate of 3.29 ms. A two-sided paired Bayesian t-test revealed moderate evidence against a difference in mean-number intercepts between the ranges and thus against AMdependency of the number mapping on the MNL with $BF_{10} = 0.10$.
Moreover, we tested whether there was a relation between the SNARC slopes in the lower and the higher range. The data revealed moderate evidence against a correlation ~~of the SNARC slopes between the lower and the higher range~~ with $BF_{10} = 0.19$."

Revised results for exploratory analyses in Experiment 2 (pages 34 and 35):

"As for Experiment 1, to disentangle the possible scenarios illustrated in Figures S1 and S2 and Table S1 (see Supplementary Material) in Experiment 2, the mean-number intercepts were compared between ranges. The mean number in the lower range from 1 to 5 (excluding 3) was 3 with a dRT estimate of 0.57 ms, and the mean number in the higher range from 4 to 8 (excluding 6) was 6 with a dRT estimate of 2.51 ms. A two-sided paired Bayesian t-test revealed moderate evidence against a difference in mean-number intercepts between the ranges and thus against AMdependency of the number mapping on the MNL with $BF_{10} = 0.18$.

As in Experiment 1, we tested whether there was a relation between the SNARC slopes in the lower and the higher range. In contrast to the data of Experiment 1, the data of Experiment 2 revealed strong evidence for a moderate correlation ~~of the SNARC slopes between the lower and the higher range~~ with an estimate of $r = 0.34$ and $BF_{10} = 521906.11$.

In contrast to Experiment 1, number parity and number magnitude were orthogonal in Experiment 2 (i.e., the mean number magnitude was equal for odd and even numbers in each range). Therefore, we were also able to test the MARC effect in Experiment 2. A two-sided Bayesian one-sample t-test revealed moderate evidence against a MARC effect in the lower range with $BF_{10} = 0.16$, and inconclusive evidence regarding a MARC effect in the higher range with $BF_{10} = 0.51$."

4. I would recomend a summary table guiding the reader to have a visual overview of the observed evidence (in terms of both tested HPs and quality of evidence).

Such a table is already included in the manuscript, following the "Study Design Table" template requested by PCI, including the observed evidence regarding each hypothesis test. We describe it on page 28:
"A Study Design Table was filled in prior to data collection and gives an overview of all research questions, corresponding hypotheses, the targeted sample size and planned analyses with a rationale, the interpretations of potential outcomes and theoretical conclusions (see Table 5). It also contains the observed outcomes for both experiments."

Additionally, we have created a new figure, which mimics Figures 3 and 4 and at the same time summarizes all findings regarding our hypotheses. We inserted it as Figure 5 to the results part of our revised manuscript:

Figure 5

Summary of all tested hypotheses and outcomes in the plot for the linear regression of mean dRTs on number magnitude separately for the lower (blue, solid line) and higher (orange, dashed line) number ranges for Experiments 1 (left panel, see Figure 3) and 2 (right panel, see Figure 4)
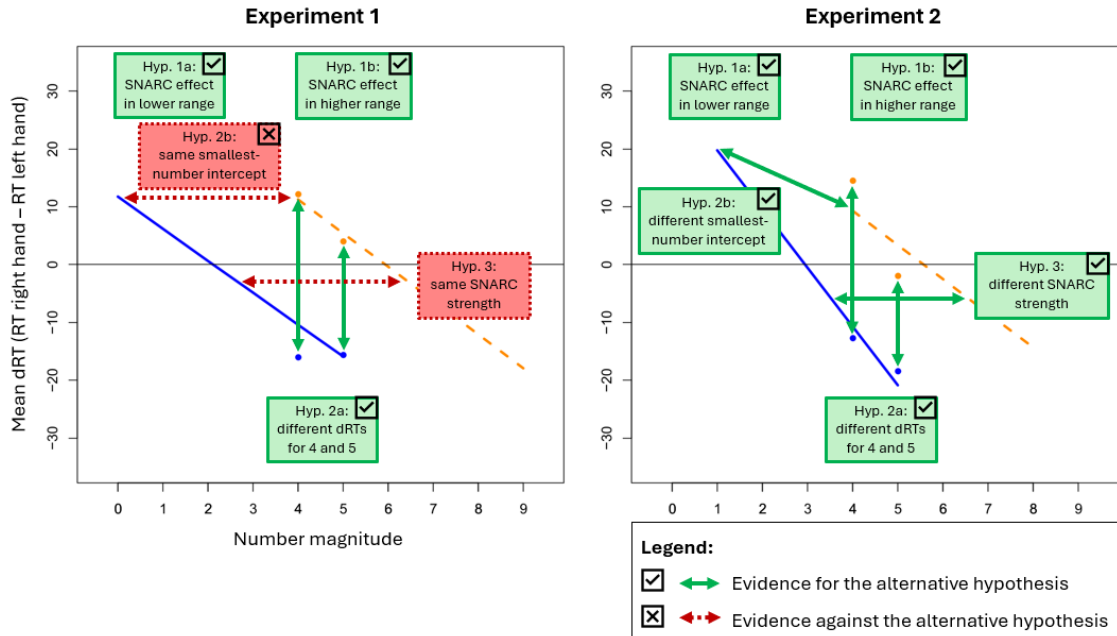
Note. The figure only includes the mean dRTs for the critical numbers 4 and 5, which appear in both the lower and the higher number ranges. Hypotheses 1a and 1b were tested with one-sample *t*-tests, whereas Hypotheses 2a, 2b, and 3 were comparisons tested with paired *t*-tests and are illustrated with two-sided arrows. Green boxes with a solid outline and a checkmark as well as green solid arrows indicate Bayesian evidence for the alternative hypothesis (i.e., $BF_{10} > 3$). Red boxes with a dotted outline and a cross as well as red dotted arrows indicate Bayesian evidence against the alternative hypothesis (i.e., $BF_{10} < 1/3$).