July 23, 2020

Study and Analysis Plan Preregistration

Study Information

Title: S-R compatibility task: Effect of stimulus size on horizontal response location – A replication study

Authors: Tim Bax (tbax@uos.de), Milan Ewert (mewert@uos.de), Niklas Bub (nbub@uos.de), Özge Özenoglu (ooezenoglu@uos.de)

Description: The *A Theory of Magnitude* (ATOM) model proposes that time, space, number and other dimensions are linked through a magnitude system within the parietal lobe (Walsh, 2003). One of the model's predictions is that there are intrinsic reciprocal interactions across magnitude dimensions.[1] If this were the case, manipulating one dimension should correspond with an interfering perception of the other dimension. Indeed, both neurophysiological and behavioral research provide evidence in favor of the ATOM model and its prediction. In behavioral research, the literature contains various studies that observed so-called SNARC effects (**S**patial-**N**umerical **A**ssociation of **R**esponse **C**odes, Dehaene et al, 1993) and size-congruency effects (e.g., Henik et al., 1982). A recent study by Wühr and Seegelke (2018) investigated yet another – rather unexplored – combination of ATOM's implications, namely whether there are compatibility effects between physical stimulus size and spatial response location. Specifically, the study results suggest that small objects are associated with the left side and large objects with the right side. As cumulative evidence in favor of the S-R compatibility effect would extend ATOM's framework, we will attempt to directly replicate the first experiment of the given study.

Hypotheses

In accordance with experiment 1 of Wühr and Seegelke (2018), we address the following research hypotheses:

I. Response times are faster in the compatible mapping condition than in the incompatible mapping condition.
II. The S-R compatibility effect is more pronounced in right-hand responses.

Besides investigating the existence of a stimulus size-response location compatibility effect as addressed with hypothesis I. above, Wühr and Seegelke (2018) also found that this effect is more pronounced in right-hand responses. Since what counts as 'more pronounced' is ambiguous, and the authors did not elaborate on their interpretation, we decided to split this hypothesis in two. We reason that if the compatibility effect is more pronounced in right-hand responses, right-hand responses should be faster in the compatible mapping condition, and slower in the incompatible mapping condition compared to left-hand responses. Therefore, the discrepancy between the reaction times through manipulation should be larger for right-hand responses than for left-hand responses. We end up with the following hypotheses of interest:

[1] Put differently, ATOM proposes that we have an internal representation of "magnitude" across axes (e.g., horizontal or vertical), where magnitude increases or decreases towards the poles of an axis. For instance, if we compare the numbers 1 and 100, we say that 1 is smaller than 100, and 100 is larger than 1. The theory further states that other dimensions (such as space, physical size, etc.), may influence the perceived magnitude of, say, the numerical size.

1. Response times are faster in the compatible mapping condition than in the incompatible mapping condition.
2. The compatible S-R mapping yields faster reaction times for right-responses than for left-responses.
3. The incompatible S-R mapping yields slower reaction times for right-responses than for left-responses.

We will conclude that the compatibility effect is more pronounced in right-hand responses only if there is compelling evidence for both hypothesis 2 and hypothesis 3 (see Inference criteria below).

Design Plan

Study type: Browser-based experiment

Blinding: The relevant manipulation, namely reversing the mapping conditions, is within-participants. In the first block (II and III, see study design below), participants are not aware that a later switch in conditions will take place. After finishing the first main test phase, participants are informed about the reversed mapping and instructed to press the keys accordingly. Participants are neither informed about the purpose of the reversed mapping conditions, nor are they informed about the objective of this study. The experiment is conducted via the internet. There will be no direct contact between participants and experimenters. The data will be analyzed by the experimenters.

Study design: The study is a within-subjects design with two factors. The two factors are mapping condition (levels: compatible, incompatible) and correct response location (levels: right, left). In each trial, participants either see a small square (2 x 2 cm) or a large square (4 x 4 cm) at the screen center. Their task is to judge whether the square is the smaller or the larger one (forced binary choice). In the *compatible* mapping condition, participants should respond to the small square by pressing the "q" key, and should respond to the large square by pressing the "p" key. In the *incompatible* mapping condition, it is the other way around: Participants should press the "q" key if they see the large square and the "p" key if they see the small square. We created the stimuli (small and large square) by ourselves. They are available here.

The experiment consists of six parts:

I. introduction & instructions
II. practice phase (first S-R mapping)
III. main test phase (first S-R mapping)

---
optional pause

IV. practice phase (second S-R mapping)
V. main test phase (second S-R mapping)
VI. post-experiment questionnaire

In the first practice phase (II), participants complete ten trials with the first S-R mapping (2 stimuli x 5 repetitions). In the second practice phase (IV), there are 20 trials to complete with the second S-R mapping (2 Stimuli x 10 repetitions). Both main phases (III and V) consist of 60 trials each (2 stimuli x 30 repetitions). The order of mapping conditions (compatible – incompatible vs. incompatible – compatible) will be randomized across participants (see Randomization below).

Please refer to the experimental design document for a more extensive description.

Randomization: The order of mapping conditions (compatible first, incompatible second block vs. incompatible first, compatible second block) will be randomly assigned to participants. The first S-R mapping condition will be announced in the instructions (I). The participants will be informed of the changed S-R mapping after the first main test phase (III). All participants see both stimuli (small and large square) throughout the experiment. Each square will be shown in a random, ad hoc defined order.

Sampling Plan

Existing data: Existing data (N = 24) comes from the study by Wühr and Seegelke (2018), which we attempt to replicate. The data do not go into the final analysis. Data from a previous pilot study (N = 4) are available and guided the specification of statistical models. These data are not included in the final analysis. Both raw data are provided in the "Links" section below. No data from the experiment to be preregistered here are available at the time of preregistration.

Data collection procedures: Participants are required to have a basic command of English to be able to follow the instructions. Of course, we cannot check if this is the case, we will thus rely on the fulfillment of this premise at participation. Furthermore, participants are required to have normal or corrected-to-normal visual acuity. Participants need to do the study on a laptop or desktop computer – phones and tablets are not allowed. These three premises will be communicated in the invitation and by participation assumed to be fulfilled. Participants are aware that they will not be reimbursed for their participation. They will be recruited via social media and email. Every participant is allowed to take part in the experiment only once. We will close the data collection three days after having sent the initial invitation (see stopping rule below).

Sample size: We aim to recruit at least 48 participants.

Sample size rationale: We aim to recruit at least twice as many participants as the original study. As the original study recruited 24 participants, we will aim to get at least 48 participants in our replication study. We reason that with a larger sample, our results are rendered more valid and our study has higher statistical power. We think that at least doubling the initial sample size will be a noteworthy increase in the sample size.

Stopping rule: Due to internal deadlines regarding the final project, we will stop data collection at 23:59 PM on the third day after sending out the first invitation.

Variables

Manipulated variables: We manipulate one variable, namely the mapping condition. This variable defines which key should be pressed by the participant if seeing either the small or the large square. It has two values: *compatible* and *incompatible*. In the compatible mapping condition, participants should respond to the small stimulus by pressing the "q" key (i.e., giving a left-hand response), and to the large stimulus by pressing the "p" key (i.e., giving a right-hand response). In the incompatible mapping condition, they should respond to the small square by giving a right-hand response, and to the large square by giving a left-hand response. Concretely, CONDITION will be treated as a 2-level factor with reference level *compatible*.

Measured variables: The single dependent variable we will measure for both confirmatory and exploratory analysis is the reaction time between stimulus onset and button press. If the time limit of 1500 ms is exceeded, or if a participant responded faster than 100 ms, we will exclude this individual trial from the analysis (see data exclusion below). Concretely, RT is a metric variable capturing the reaction times. Furthermore, we store the mapping condition in the binary variable CONDITION (values: compatible, incompatible). This variable will be manipulated by us (see also Manipulated variables above). We will further measure whether each trial was correct or not. CORRECTNESS is a binary variable with values *correct* or *incorrect*. We will use this variable to filter out incorrect trials. After filtering out incorrect trials, we will create a binary variable CORRECT_RESPONSE, which has the values *left* or *right*. We create this variable based on the values of binary variables CONDITION (compatible, incompatible) and EXPECTED (small, big). The variable RT will be used as the variate, and CONDITION and CORRECT_RESPONSE will serve as covariates (see analysis plan below).

Another variable of interest is the nominal variable HANDEDNESS. Initially, the values are stored in variable EDU (for 'education') in the post-experiment questionnaire. However, we changed the predefined default question about the participant's education level in _magpie. The reason for this is that changing the presented question of a default variable in the predefined post-questionnaire function is much easier to implement than defining a custom-made post-questionnaire function that includes the desired HANDEDNESS variable. So, instead of giving optional information about the education level (which is not a variable of interest anyway), participants are now encouraged to indicate their dominant hand (options: left, right, I am able to use both hands equally well). In the data wrangling stage, we will immediately rename the measured EDU variable to HANDEDNESS. HANDEDNESS is a variable that we are interested in for the exploratory analysis (see the section on Exploratory Analysis below).

Indices: We do not consider any indices.

Analysis Plan

Statistical models: We will fit one Bayesian hierarchical model with variate RT (the reaction time) and covariates CONDITION (compatible, incompatible), CORRECT_RESPONSE (right, left), and their two-way interaction. We add by-participant random intercepts and slopes for both covariates. We will address all three hypotheses by referring to this model. We chose this model after comparing various models using the Leave-One-Out Cross-Validation (LOO-CV) method. Please refer to the model comparison script for our justification. We will analyze the data with the statistical programming language R. We will rely on the Stan modeling language (Carpenter et al., 2016) through the `brms` package (Buerkner, 2016) and use the `faintr` package introduced in Franke and Röttger (2019) for hypothesis testing. The analysis script 02-SR-compatibility-analysis-pilot.html contains the analysis as planned at the time of handing it in for feedback (as of July 15, 2020).[2]

[2] After obtaining feedback on our study and analysis plan preregistration, we edited our analysis script for the main experiment accordingly. The revised analysis script is available here: https://github.com/ooezenoglu/XPlab-2020-SR-Compatibility/blob/master/analysis/02_main/01-SR-compatibility-analysis-main.Rmd

Note: At the time of revising the analysis script for the main experiment, we did not have any data from the main experiment to be preregistered here. Therefore, this file may potentially not contain any data from the main experiment if checked while data collection still is taking place.

Transformations: Not transformations will be applied. The categorical variables CONDITION and CORRECT_RESPONSE will use treatment coding with reference levels *compatible* and *right*, respectively.

Inference criteria: We will use a posteriori credible values for the effect coefficients CONDITION and CORRECT_RESPONSE. We will rely on the `compare_groups` function of the `faintr` package. We judge there to be a credible effect of the manipulation *compatible mapping* vs. *incompatible mapping* (hypothesis 1) if the 95% credible interval for the mean difference of the posterior for a group comparison of 'incompatible' (higher) and 'compatible' (lower) is entirely positive and does not include zero. We will conclude that S-R compatibility effects are more pronounced in right-hand responses only if there is compelling evidence for both hypotheses 2 and 3. We confirm hypothesis 2 if the 95% credible interval for the mean difference of the posterior for a group comparison of 'left' (higher) and 'right' (lower) is greater than zero for the 'compatible' condition. We will confirm hypothesis 3 only if the 95% credible interval for the mean difference of the posterior for a group comparison of 'right' (higher) and 'left' (lower) is greater than zero for the 'incompatible' condition.

Data exclusion: In accordance with the original study by Wühr and Seegelke (2018), we exclude every individual trial faster than 100 ms and slower than 1500 ms. As we are only dealing with reaction times and not error rates, we will use the binary variable CORRECTNESS (correct, incorrect) to filter out incorrect trials. No data from practice trials will enter the analysis.

Missing data: Should a data set not be recorded completely, we will use all data available from that participant.

Exploratory Analysis: Wühr and Seegelke (2018) noted in the "Conclusion and directions for future research" section that the "participant's handedness may modulate the effect as well" (p. 9). As their study results are only based on data from right-handed participants (N = 24), the suggested internal representation of stimulus size and horizontal location may differ for left-handers. We therefore want to additionally investigate whether handedness modulates the S-R compatibility effect in a sense, that the internal horizontal representation might be switched. Concretely, we are interested in whether right-handers associate large objects with the right side, and small objects with the left side, and vice versa for left-handers. Questions 4 and 5 address this theory of ours. We think that a positive outcome for questions 4 and 5 may inspire follow-up confirmatory research that investigates this relationship and ultimately may shed more light on the internal representation of magnitude.

It is further reasonable to think that – regardless of the stimulus size – left-handers respond faster with the left hand, and vice versa for right-handers. Therefore, in question 6 below, we would like to test whether, in both conditions, right-handers were faster in right-responses. Conversely, in question 7, we would also like to test whether, in both conditions, left-handers were faster for left-responses.

If left-handed participants take part in our replication study (and indicated their left-handedness), we would like to address the following questions:

4. Do right-handers respond faster in the compatible mapping condition than left-handers?
5. Do left-handers respond faster in the incompatible mapping condition than right-handers?
6. Do left-handers respond faster to left-responses than to right-responses?
7. Do right-handers respond faster to right-responses than to left-responses?

As both-handedness is very rare, we assume that there will probably be no participant that indicates both-handedness. If, however, there are data from both-handed participants (via self-report), we are happy to also further explore that data.

Tentative Analysis Plan for Exploratory Analysis: We plan on running a Bayesian hierarchical model with variate RT and covariates CONDITION, CORRECT_RESPONSE and HANDEDNESS (with three-way interaction). Furthermore, we specify by-participant random intercepts and slopes for the interaction of covariates CONDITION and CORRECT_RESPONSE. For testing the abovementioned questions, we will again use the `compare_groups` function of the `faintr` package.

Links

GitHub repository:
https://github.com/ooezenoglu/XPlab-2020-SR-Compatibility

Materials:
https://github.com/ooezenoglu/XPlab-2020-SR-Compatibility/tree/master/experiment/02_main/materials

Experimental design:
https://github.com/ooezenoglu/XPlab-2020-SR-Compatibility/blob/master/writing/01-experimental-design-SR-Compatibility-Task.pdf

Link to raw data (N = 24) provided by Wühr and Seegelke (2018):
https://doi.org/10.5334/joc.19.s1

Raw data of the pilot study (N = 4) by us:
https://raw.githubusercontent.com/ooezenoglu/XPlab-2020-SR-Compatibility/master/data/01_pilot/01-raw-data-pilot.csv

HTML preview Model Comparison:
https://htmlpreview.github.io/?https://github.com/ooezenoglu/XPlab-2020-SR-Compatibility/blob/master/analysis/01_pilot/01-SR-compatibility-model-comparison.html

HTML preview Analysis Script (pilot):
https://htmlpreview.github.io/?https://github.com/ooezenoglu/XPlab-2020-SR-Compatibility/blob/master/analysis/01_pilot/02-SR-compatibility-analysis-pilot.html

Revised Analysis Script after Feedback (main):
https://github.com/ooezenoglu/XPlab-2020-SR-Compatibility/blob/master/analysis/02_main/01-SR-compatibility-analysis-main.Rmd

References

Buerkner, Paul-Christian (2016). "brms: An R package for Bayesian multilevel models using Stan". In: Journal of Statistical Software 80.1, pp. 1–28.

Carpenter, Bob et al. (2016). "Stan: A probabilistic programming language". In: Journal of Statistical Software 20, pp. 1–37.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122, pp. 371–396. DOI: https://doi.org/10.1037/0096-3445.122.3.371

Franke, M., & Roettger, T. B. (2019). Bayesian regression modeling (for factorial designs): A tutorial. DOI: https://doi.org/10.31234/osf.io/cdxv3

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, 10, pp. 389–395. DOI: https://doi.org/10.3758/BF03202431

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends In Cognitive Sciences*, 7, pp. 483–488. DOI: https://doi.org/10.1016/j.tics.2003.09.002

Wühr, P., & Seegelke, C. (2018). Compatibility between Physical Stimulus Size and Left-right Responses: Small is Left and Large is Right. *Journal of Cognition*, 1(1): 17, pp. 1–11. DOI: http://doi.org/10.5334/joc.19