July 14, 2020

Study and Analysis Plan Preregistration

Study Information

Title: S-R-compatibility task: Effect of stimulus size on horizontal response – 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. 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, Dahaene 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:

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

Design Plan

Study type: Browser-based experiment

Blinding: The relevant manipulation, namely reversing the mapping condition, is within-participants. In the first block (II and III, see study design below), participants are not aware that a later switch in condition will take place. After finishing the first main test phase, participants are informed about the reversed mapping and instructed to press the keys accordingly. 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 10 trials with the first S-R mapping (2 stimuli x 5 repetitions). In the second practice phase (IV), there are twenty 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 below).

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

Randomization: The order of mapping conditions (compatible first, incompatible second block or 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 reminded 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 random, ad hoc 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 good command of English to follow the instructions. 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 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 10 days after having sent the initial invitation (see stopping rule below).

Sample size: We aim to recruit at least 60 participants.

Sample size rationale: To assure at least 80% power, we follow the proposal of Simonsohn (2015) and calculate the minimum sample size with:

$$replication\ sample\ size\ =\ 2.5 \times original\ sample\ size\ =\ 2.5 \times 24\ =\ 60$$

Stopping rule: Due to internal deadlines regarding the final project, we will stop data collection at midnight on the 10th day after sending out the first invitation. If, however, we achieved the minimum sample size of 60 participants before this deadline, we will stop data collection right after obtaining the data of the 60th participant.

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 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 (compatible, incompatible). We will further measure whether each trial was correct or not. CORRECTNESS is a binary variable with value *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).

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. 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.

Transformations: Not transformations will be applied. 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, and their two-way interaction. We judge there to be a credible effect of the manipulation *compatible mapping* vs. *incompatible mapping* (hypothesis 1) if the 95% credible interval of the posterior for the `conditionincompatible` and `conditionincompatible:correct_responseleft` coefficients only contain values greater than zero. Furthermore, the posterior probability of the difference δ being bigger than zero should be close to zero. Concretely, we will check if P('higher – lower' > 0) is close to zero, where 'higher' will be the `condition:compatible` group, and 'lower' will be the `condition:incompatible` group.

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 of the posterior for the `correct_responseleft` coefficient only contains values greater than zero. Further, P('higher – lower' > 0) should be close to zero, where 'higher' is the right response for the compatible condition, and 'lower' is the left response for the compatible condition. We will confirm hypothesis 3 only if the posterior probability of δ being bigger than zero is close to zero. Concretely, P('higher – lower' > 0) should be close to zero, where 'higher' is the left response for the incompatible condition, and 'lower' is the right response 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. 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: Furthermore, we are interested in the following questions:

1. In the compatible condition, left-handers respond faster to left responses than to right responses.
2. In the incompatible condition, left-handers respond faster to right responses than to left responses.
3. In the compatible condition, right-handers respond faster to right responses than to left responses.
4. In the incompatible condition, right-handers respond faster to left responses than to right responses.

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

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

Simonsohn, U. (2015). Small Telescopes: Small Telescopes: Detectability and the Evaluation of Replication Results. *Psychol Science*, 26(5): 559–569. DOI: https://doi.org/10.1177/0956797614567341

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