Study Information

Title: Perception and Identification of Randomness

Authors: Linus kleine Kruthaup, Cosima Oprotkowitz, Mara Rehmer, Katharina Trant

Description: Perception of randomness, the ability to perceive and discriminate structured versus unstructured events, is an important ability for survival and involved in many day-to-day activities and thus also important to study.

In this replication study, discrimination and identification of random versus non-random stimuli are tested between-participants in two experimental groups. For an exploratory analysis, we are looking into a possible effect of language on the identification of randomness, as the experiment is conducted in German and in English.

Hypothesis: "The probability of correctly identifying stimuli from R [random sources] and N [non-random sources] coincides with the ease of distinguishing between the two sources." (Zhao & Hahn, 2014)

Design Plan

Study type: Experiment

Blinding: The participants are not informed that there are two different experimental groups and each participant is only assigned to one group. The experiment will be conducted via the internet, so no direct contact between experimenters and participants will take place.

Study design: The experiment is a 2x51 mixed-factorial design. The first factor is *condition*, with the two values *discrimination* and *identification*. The second factor is *switch rate*, which has 51 levels. Each participant contributes data points for only one *condition* (between-participants), but multiple data points (ideally 10) for *switch rate* (within-subject, repeated measures).

The overall design is the same as in the replicated experiment, but we made some changes to the colors of the stimuli materials and added the instructions in German.

A more detailed description can be found in the corresponding Experimental Design plan.

Randomization: The participants are randomly assigned to either the discrimination or the identification condition by a coin-flip generator.

Sampling Plan

Existing data: As of the date of the submission of this preregistration, the data have not yet been collected.

Data collection procedures: Participants will be drafted through social media and direct messages (e-mails and text messages). Participation is voluntary and will not be compensated. After sending out the invitations on the 1st of August, we will wait 12 days until closing the data collection. Participants are only eligible if they are at least 18 years old and, although this may seem redundant, if they have full or corrected vision. A participant is only allowed to participate once.

Sample size: We will try to recruit as many participants as possible.

Sample size rationale: Since time is critical due to a deadline, our pool of reachable participants is limited and we do not offer any compensation for participation, we cannot state a minimum number of participants.

Stopping rule: We will stop data collection on noon of the 12th of August, which is 12 days after sending out the invitations, due to the project deadline.

Variables

Manipulated variables: Firstly, we manipulate the condition, as some participants are assigned to one and other participants to the other condition. We then manipulate the switch rate, that is the density of randomness, of the given stimuli picture. The switch rate has 51 different levels, each of which is shown ten times per participant. This switch rate lies between 0 and 1. The closer to 0.5 the switch rate, the more does the color assignment happen like one would expect from a random source. The manipulated variable *switch rate* is the same for discrimination and identification. More on this in the 'Experimental design'.

Measured variables and indices: We will measure the reaction time at each task (for data exclusion, see below) and we will measure whether the tasks were answered correctly or not. Then we will calculate the average accuracy, that is the proportion of correctly answered tasks, at every switch rate for the discrimination and the identification condition. We will measure this by first calculating the average accuracy at every switch rate for each participant and then calculate grand means by averaging across the participants in the respective condition.

Analysis Plan

Statistical models: We will have a Bayesian linear regression model, where *accuracy* is the dependent variable and *condition* (group) and *switch rate* are independent variables. We will conduct the analysis using the programming language R and the ´brms´ package. We will use the (default) flat priors of ´brms´ and the formula we will use for our model is *accuracy ~ condition * switch rate.*

This is still preliminary, because we are hoping for helpful feedback here.

The script "Statistical-Analysis" from our pilot contains our analysis as planned.

1. Inference criteria (optional)

    1.1. *What criteria will you use to make inferences? Please describe the information you'll use (e.g. p-values, bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?*

    1.2. ***Example****: We will use the standard p<.05 criteria for determining if the ANOVA and the post hoc test suggest that the results are significantly different from those expected if the null hypothesis were correct. The post-hoc Tukey-Kramer test adjusts for multiple comparisons.*

    1.3. ***More information:*** *P-values, confidence intervals, and effect sizes are standard means for making an inference, and any level is acceptable, though some criteria must be specified in this or previous fields. Bayesian analyses should specify a Bayes factor or a credible interval. If you are selecting models, then how will you determine the relative quality of each? In regards to multiple comparisons, this is a question with few "wrong" answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.*

Data exclusion: We will exclude data points with reaction times slower than 8000ms, as we cannot guarantee that the answer is related to the respective stimulus when that much time has passed.

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

Exploratory analysis: We further plan to look for relationships between the language the participants stated as their main language and the results of the identification part. (This of course only applies to those in the identification group.)

Other

This study is a replication of Experiment 1 in "Perception and identification of random events" by Zhao and Hahn (2014) (https://doi.org/10.1037/a0036816). Changes we made to their experimental design and analysis are either stated here or in our 'Experimental design'.