Discriminating and Identifying Randomness

A tool for making sense of the world

Abstract

Discriminating randomness from structure and identifying a random pattern as such are important abilities for survival. In this replication study we investigate these skills with the use of binary colored matrices. The results indicate that discrimination is less difficult than identification, although we are not strongly confident in these results due to some impediments.

Keywords: Randomness, Identification, Discrimination, Replication

Introduction

Humans tend to make sense of the world by finding structure or patterns. Thus, the ability to perceive and process the lack thereof, randomness, is crucial and has been subject to quite some studies (see Bar-Hillel & Wagenaar, 1991), even though there is no clear or definite definition of randomness. Being able to identify structure and patterns as well as being able to generate random behavior is important for survival. Prior studies about the perception of randomness have shown that the human understanding of randomness can be biased: The gambler's fallacy for example describes the belief that an event is less likely to occur in the future if it has occurred more frequently in the past (or vice versa), although the probability of such events does not depend on what has happened in the past (Kahneman & Tversky, 1972). While other studies use judgment of random sequences and generation of random sequences tasks (for example Warren et al., 2018), we replicate 'Experiment 1' of Zhao et al. (2014). In this experiment we compare performance in a *discrimination* versus an *identification* task to understand the conceptualization of randomness.

Hypothesis

Having a correct and concrete conceptualization of randomness should lead to high accuracy in discriminating a border between randomly and non-randomly generated stimuli. Likewise, this conceptualization should also lead to similar performance in a task asking for identification of a randomly generated stimulus. So, we apply the hypothesis used in the original paper (Zhao et al., 2014):

"*Hypothesis 1*: The probability of correctly identifying stimuli from R and N coincides with the ease of distinguishing between the two sources."
Where R is a random source and N is a non-random source.

Method

Participants

We had a total of 19 participants who took part in our version of the 'Perception of Randomness' task. There were 5 male, 12 female and 2 participants who chose not to state any gender, with ages ranging from 20 to 64. Their participation was voluntary and not compensated.
10 participants took part in the discrimination condition and 9 in the indentification condition.
The first 4 participants contributed 510 data points each (10 for each switch rate), the subsequent participants contributed 306 data points each (6 for each switch rate) plus 10 practice trials.

Materials
Each stimulus consists of a 60 × 60 pixel matrix. The matrix is either separated horizontally or vertically in half, for each stimulus determined at random. Each cell of

the matrix can have either one of two colors. In the original paper these two colors were green and blue, we changed it to visually impaired friendly blue and orange. While one half of the matrix is generated at random, the other half is generated using a stochastic algorithm. In this algorithm which is called "$switch(x)$", $x \in [0,1]$ is the switch rate which determines the probability that the $n + 1^{th}$ cell is set to the opposite value of the nth cell. In all cases, the color of the first cell is chosen randomly. For a switch rate of $x = 1$ the cells perfectly alternate in color, for a switch rate of $x = 0$ the cells all have the same color. In both cases the outcome is determined by the first cell. A switch rate of $x = 0.5$ behaves like a fair coin and assigns one of the two colors at random to a cell. A switch rate of $x < 0.5$ or $x > 0.5$ has longer or shorter runs than expected from a random source, respectively. In both conditions the switch rate of the non-random half is between 0 and 1 in steps of 0.02, consequently leading to 51 levels. Each level is repeated 6 times[1], so that we end up with 306 trials. The difference in stimulus between the two experimental conditions *discrimination* and *identification* is that for the *identification* task there is a 50-pixel gap horizontally or vertically oriented showing the separation of the two halves.

The whole experiment was implemented in _magpie, a minimal architecture for the portable generation of interactive experiments, written in CSS, Javascript and HTML. The "$switch(x)$" algorithm described above is implemented using the Javascript function Math.random(), a function that generates a pseudorandom floating number between 0 and 1 (1 exclusive). The sequence generated by the " $switch(x)$" algorithm is then used to insert the respective colors pixel-wise into a canvas object, which is then enlarged by a factor of 3 and transformed into an image.

In the *discrimination* condition, the participants are shown matrices without separation gap as stimuli. Their task is then to discriminate the two halves by indicating whether a separating line should be placed horizontally or vertically by pressing the *'v' key* for vertical and the *'h' key* for horizontal. Instructions of the discrimination condition do not include mentioning of randomness or probability.

In the *identification* condition, the presented stimuli are matrices with a horizontally or vertically oriented gap, separating the random half from the one with the non-random switch rate. The task is then to decide which of the two halves was more likely generated by a random process. This is done by pressing *'w' key* (up) or *'s' key* (down) if the division is horizontal, and *'a' key* (left) or *'d' key* (right) if vertical.

Procedure

The whole experiment took place online and consisted of 4 parts:
1. introduction and instructions
2. practice trials
3. test phase
4. post-experiment questionnaire

The participants were first blindly assigned to one of two different experimental groups via an implemented coin flip and then welcomed to the experiment. They received different instructions depending on their experimental condition as specified below. Before the main experiment, the participants were presented with 10 practice trials to become familiar with the task. In the original paper, practice trials were not mentioned explicitly, but we considered the tasks to be quite challenging and in need of some practice beforehand. In both cases, stimuli were shown for 1500 ms followed by a blank screen until a response

---

[1] The first 4 participants contributed 510 data points (10 for each switch rate)

was entered. The order of the stimuli was shuffled before each participant's trials.

The original instructions were altered for both conditions to better fit our experiment conditions and can be read down below.

Discrimination:

'In each trial of the experiment you will see a square made up of two colours. Each matrix can be divided into two halves either horizontally (-) or vertically (|). The two halves are generated from different processes. Your task is to judge the orientation of the boundary between the two halves, by pressing 'v' key for vertical or 'h' key for horizontal. Please press the respective key when the picture is gone. Before we start there will be 10 practice trials, so you can familiarize yourself with the task. We will tell you exactly when the main experiment starts.'

Identification:

'In each trial of the experiment you will see a square made up of two colors. Each matrix is divided into two halves either horizontally (-) or vertically (|). The two halves are separated by a gap. One halve is generated from a random process and the other from a nonrandom process. Your task is to identify which half is more likely to be produced by a random process than a nonrandom process. Press the 'w' key (up) or 's' key (down) if the division is horizontal, and 'a' key (left) or 'd' key (right) if vertical. Please press the respective key when the picture is gone. Before we start there will be 10 practice trials, so you can familiarize yourself with the task. We will tell you exactly when the main experiment starts.'

There was no feedback regarding the correctness of the responses. After the experiment was finished the participants had the chance to leave additional

sociodemographic information (age, gender, native language, level of education) and feedback. Furthermore, they were asked whether they have an idea what the experiment is about in an open question.

The whole experiment was provided in English and German.[2]

Design

The experiment is a 2 x 51 mixed factorial design, where the first factor is *condtion* with two values *discrimination* or *identification*. The second factor is *switch_rate* which has 51 levels (from 0 to 1 in steps of 0.02). Each participant contributes data points for only one condition (between-subjects) but multiple data points (6 or 10) for each switch rate (within-subject, repeated measures).

Data preparation

We chose to exclude individual trials with a reaction time bigger than 8000 ms, since we can not guarantee that participants recall the stimulus correctly after this time.
There was no exclusion for too fast responses as answers could only be entered after the stimulus was shown for 1500 ms.
This resulted in the exclusion of 73 trials.

Results

Analysis
For our analysis we used a Bayesian Logistic Regression Model which models the *correctness* as the dependent variable in terms of *condition* and *switch_rate* as independent variables. We chose this measure of proportion of correct responses in each condition and switch rate instead of, like in the original paper, calculating summary statistics like accuracy for the model. Furthermore, we decided not to average over all participants and switch rates in the two conditions since the

---

[2] German instructions can be found in the Appendix

summary of the data will lead our model to underestimate the variance in the data.

Additionally, we included random slopes for each participant because the model assumes that the data points are independent of each other, yet each participant contributes data points for each switch rate.

We used weakly informative priors for all *class b* parameters, namely *student_t(1,0,2)* prior distribution since they speed up the sampling. We ran 5000 iterations and set a seed (13).

Results

We first looked at the summary statistics like the mean accuracy for both conditions to get an overview of our data. Straight away we noticed some discrepancies to prior expectations. The mean accuracy for the *discrimination* task lies at 0.765, which indicates that a majority of trials were answered correctly on average, whereas accuracy lies at 0.507 for the *identification* task, which would indicate that responses were almost given by chance.

We expected that there are more correct responses than incorrect responses, especially for low and high switch rates, as it should get easier to identify which half of the shown matrix was generated at random or discriminate whether the border between the two halves is vertical or horizontal.

To have a thorough look at the data, we plotted each participant's responses in each switch rate[3].

There were 4 performances, mostly in the *identification* condition, that looked like the participants might have misunderstood the task to always choose the 'randomly generated' matrix but chose the one that did not look random - although we can only assume this. Only one performance in the *discrimination* task seemed unexpected and as if they confused 'vertical' with 'horizontal'. Three performances lead us to think that the

participants might not have completely understood the task as there is no clear pattern to be seen in their responses.

Unfortunately, this is not the best data to work with, as we do not have enough data to compensate for these discrepancies. Since we only chose to exclude trials with too long reaction times, with no intentions to exclude participants due to lack of performance, and we can not simply invert the performance of some participants where we implied that the data looked upside down, we assume that the data we measured is representative of performance in such tasks and continued with our analysis as planned although the deviation already weakens our confidence in the results.

For the inference criteria we use 95% credible intervals (CI) from the posterior estimates and we will only be confident in these estimates if their 95% CI does not include zero.

The base level estimate (*Intercept*) for switch rate = 0 in the discrimination condition is 1.61, corresponding to a proportion of correct answers of approximately 0.83, the 95% CI is quite large [0.67, 2.52] but only includes positive numbers. The estimate for *conditionidentification* is -1.39 and the 95% CI ranges from [-2.74, -0.05], which is why we are confident that the proportion of correct answers is less in the identification condition.

There are only some levels of the *swr* (switch rate) and the interaction terms between *conditionidentification:swr* that are meaningful since their CI does not include zero. The estimates for the switch rate = 0.4 to switch rate = 0.66 are negative which indicates that the proportion of correct answers goes down. This is plausible since we expected it to be more difficult for switch rates closer to 0.5 to tell apart the random from the non-random half in both conditions.

For the switch rates = 0.36 to switch rates = 0.66 the interaction estimates are credible

---

[3] the plot can be found in the Appendix: Figure 2

Posterior estimated proportion of correct answers across switch rates for both conditions

With original data points and 95% Credible Intervals

Figure 1

above zero, meaning that there were more mistakes made in the discrimination condition for these switch rates, which coincides with our observation.

There is a substantial variability across the participants (*submission_id*) with an estimate of 1.43 which is also consistent with our observations of high variability in performance.

The plot in Fig. 1 shows the posterior estimates and their 95% CI for the proportion of correct answers in each switch rate. The visualization of the discrimination condition looks as expected. For lower and higher switch rates (further away from 0.5) the estimated performance is quite high (above 0.75). At switch rates around 0.3 the proportion of correct answers goes down, even below 0.5, and goes up again until a switch rate of around 0.74 where the performance stabilizes at 0.85 accuracy.

The plot of the *identification* condition on the other hand does not look like we expected. Here we only see a jagged line that stays around 0.5 which is just chance. The proportion of correct answers do not indicate some kind of trend, they seem to be evenly distributed.

Following this, we can assume that there is an obvious difference between *discrimination* and *identification* of random stimuli.

Conclusion

In reference to the initial hypothesis, we cannot draw a strongly confident conclusion, because we only have little confidence in the collected data, as described above. What the data suggests, nevertheless, is a definite difference between the conditions, indicating that the probability of identifying random from non-random stimuli does not coincide with the probability of discriminating between random and non-random stimuli. The data rather suggests that identification of randomness - in this context - is indeed more difficult than the discrimination. While this contradicts the initial hypothesis, these findings align with the findings in the original paper (Zhao et al., 2014).

For further research, we would recommend conducting the experiment in a controlled setting, for example a laboratory, to avoid any effects relating to the surroundings, equipment or possible distractions. Besides these possible effects, our study was admittedly limited by the number of participants, which means that excluding data points has a bigger impact on the overall data points and that we cannot be strongly confident about the results due to the small sample size. A bigger sample size would have allowed for more possible exclusions of

outliers and would have provided more data, enabling a more meaningful analysis.

Something that we did not implement was feedback on the correctness of the answers, not even in the practice trials, as this was not done in the original experiment. However, Zhao et al. (2014) conducted a follow-up study that suggested that feedback significantly improves performance, at least in the *identification* task, so we can assume that feedback would have increased the amount of correctly answered trials.

Initially, we had thought to conduct an analysis exploring the effects of language on the *identification* of randomness, as randomness is explicitly mentioned in the instructions for this condition and thus presupposes a conceptualization in the participants. One such concept, where research found that language makes a difference in the perception and conceptualization, is time (Lera, 2001).

Unfortunately, our data did not allow for a significant analysis of language effects, as most participants stated German as their native language and the experiment was not implemented in a way that would allow for a clear distinction as to in which language the participants read the instructions. Here, a follow-up study on the *identification* condition, specifically designed to look into language effects, could give insight into how language shapes our concepts of randomness and show whether for example the difference between 'Zufälligkeit' and 'randomness' indeed only lies in the difference in language. For future research, we can further remark that researching the perception and identification of randomness is far from irrelevant and that there are surely multiple methods which can be applied to look further into how humans or other animals interact with randomness. These interactions are of course not limited to different contexts of perception and identification, but also include the conceptualization and production of

randomness as well as the biases affecting these processes (for further examples see: Bar-Hillel & Wagenaar, 1991; Zhao, 2013; Warren et al., 2018) - to list a few aspects.

Such investigations hold the power to shed further light on how we make sense of the - maybe not so - structured world surrounding us.

Bibliography

Bar-Hillel, M., & Wagenaar, W. (1991). The Perception of Randomness. *Advances in Applied Neuroscience*, *12*, 428-454. 10.1016/0196-8858(91)90029-I

Franke, M. (2019). *An Introduction to Data Analysis*. https://michael-franke.github.io/intro-data-analysis/index.html

Franke, M., & Roettger, T. B. (2019, July 13). *Bayesian regression modeling (for factorial designs): A tutorial*. 10.31234/osf.io/cdxv3

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3(3)*, 430–454. 10.1016/0010-0285(72)90016-3

Lera, B. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, *43*, 1-22. 10.1006/cogp.2001.0748

_magpie. (n.d.). *_magpie*. _magpie. https://magpie-ea.github.io/magpie-site/

Warren, P., Gostoli, U., Farmer, G. D., El-Deredy, W., & Hahn, U. (2018). A Re-Examination of "Bias" in Human Randomness Perception. *Journal of Experimental Psychology: Human Perception and Performance*, *44*, 663-680. 10.1037/xhp0000462

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R.,* 232-244. Routledge. 10.4324/9781315165547

Zhao, J. (2013). *Randomness Cognition: Linking Perception and Conception*. Princeton University. http://arks.princeton.edu/ark:/88435/dsp01mg74qm18h

Zhao, J., Hahn, U., & Osherson, D. (2014). Perception and identification of random events. *Journal of Experimental Psychology: Human Perception and Performance*, *40(4)*, 1358–1371. 10.1037/a0036816

<center>**Appendix**</center>

[2] **German Instructions**

Discrimination condition:

'In jedem Durchgang des Experiments wirst du ein zweifarbiges Viereck sehen.Jede Matrix kann horizontal (-) oder vertikal (|) in zwei Hälften geteilt werden. Die beiden Hälften werden aus unterschiedlichen Prozessen erzeugt. Deine Aufgabe besteht darin, die Ausrichtung der Grenze zwischen den beiden Hälften zu beurteilen, indem du die 'v' Taste für vertikal oder die h' Taste für horizontal drücken. Bitte drücke erst eine Taste, wenn das Bild nicht mehr zu sehen ist. Um dich mit der Aufgabe vertraut zu machen beginnen wir mit 10 Übungsdurchläufen. Wir werden dir genau mitteilen wann das Hauptexperiment startet.'

Identification condition:

'In jedem Durchgang des Experiments wirst du ein zweifarbiges Viereck sehen. Jede Matrix ist entweder horizontal (-) oder vertikal (|) in zwei Hälften geteilt. Die beiden Hälften sind durch einen Spalt getrennt. Eine Hälfte wird aus einem zufälligen Prozess und die andere aus einem nicht-zufälligen Prozess erzeugt. Deine Aufgabe besteht darin, zu identifizieren, welche Hälfte eher durch einen zufälligen Prozess erzeugt wird als durch einen nicht-zufälligen Prozess. Drücke die w' Taste (oben) oder 's' Taste (unten), wenn die Teilung horizontal ist, und nach 'a' Taste (links) oder 'd' Taste (rechts), wenn sie vertikal ist. Bitte drücke erst eine Taste, wenn das Bild nicht mehr zu sehen ist. Um dich mit der Aufgabe vertraut zu machen beginnen wir mit 10 Übungsdurchläufen. Wir werden dir genau mitteilen wann das Hauptexperiment startet.'

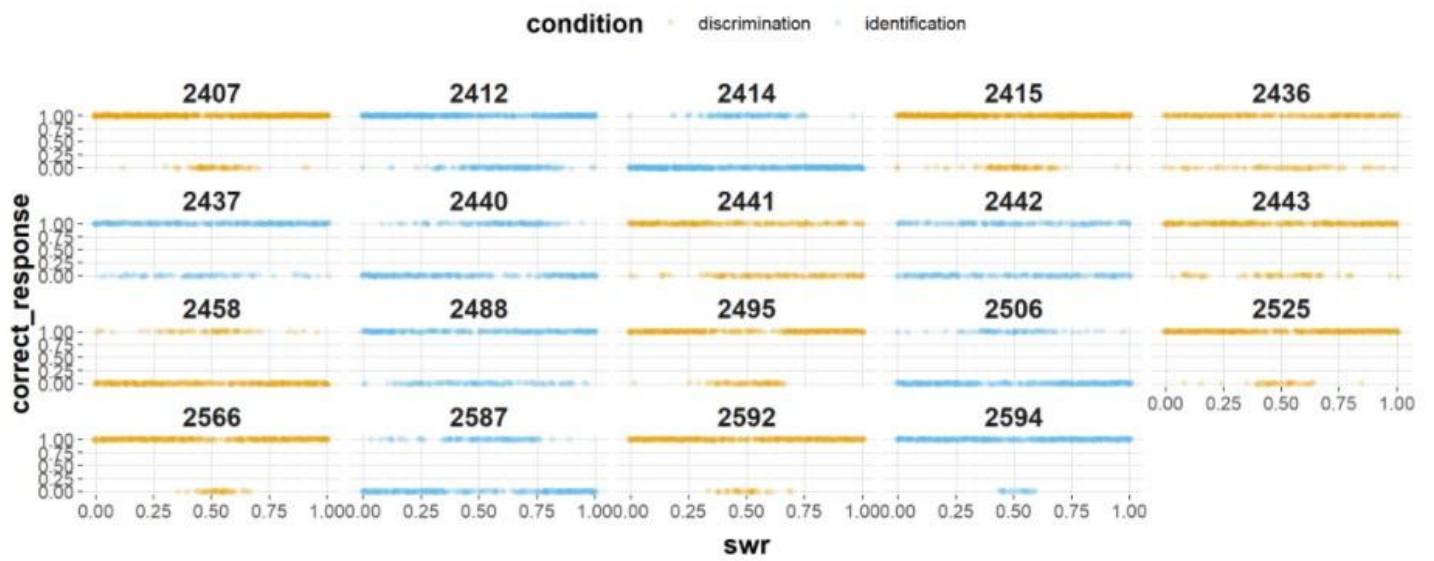**3 Plot of correct answers based on submission ID**

Figure 2