

Central Tendency as Consequence of Experimental Protocol

Stefan Glasauer (stefan.glasauer@b-tu.de)

Computational Neuroscience, Brandenburg University of Technology Cottbus-Senftenberg, Universitätsplatz 1
01968 Senftenberg, Germany

Zhuanghua Shi (strongway@psy.lmu.de)

Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13
80802 München, Germany

Abstract:

Perceptual biases found experimentally are often taken to indicate that we should be cautious about the veridicality of our perception in everyday life. Here we show, to the contrary, that such biases may be a consequence of the experimental protocol that cannot be generalized to other situations. We show that the central tendency, an overestimation of small magnitudes and underestimation of large ones, strongly depends on stimulus order. If the same set of stimuli is, rather than being presented in the usual randomized order, is applied in an order that displays only small changes from one trial to the next, the central tendency decreases significantly. This decrease is predicted by a probabilistic model that assumes iterative trial-wise updating of a prior of the stimulus distribution. We conclude that the commonly used randomization of stimuli introduces systematic perceptual biases that may not be relevant in everyday life.

Keywords: time perception, serial dependence, Bayesian estimation, iterative updating, perceptual bias

Introduction

The “central tendency” (Hollingworth, 1910) is a perceptual bias affecting estimation of magnitudes such as distance, duration, loudness, brightness, etc.: large magnitudes are underestimated, while small magnitudes are overestimated. It was first described by Vierordt (1868) for duration reproduction and has since then been re-described and rediscovered many times (see Glasauer & Shi, 2018). Until a few years ago, the findings of Vierordt have been described as an unexplained problem that “currently defies any coherent theoretical treatment” (Lejeune & Wearden, 2009). Even though the first study offering a quantitative probabilistic theory explaining the central tendency (Laming, 1999) was apparently overlooked by the scientific community, during the last few years several studies explained the central tendency as result of Bayesian estimation or similar approaches (Jazayeri & Shadlen, 2010; Petzschner & Glasauer, 2011; Bausenhardt, Dyjas & Ulrich 2014; see Shi, Church & Meck, 2013, for review).

However, the consequences of the underlying models differ to a certain extent. The model by Jazayeri & Shadlen (2010) assumes a static prior distribution (after an initial training phase). The model by Petzschner & Glasauer (2011) proposes that the prior distribution is iteratively updated from trial to trial. Therefore, the iterative updating model in contrast to the static model predicts 1) serial dependency, that is, the error in trial k depends on the stimulus difference between trial k and $k-1$, and 2) that the strength of the central tendency depends on the order of stimuli. Both consequences are coupled: if the error increases with increasing stimulus difference, then keeping differences as small as possible will minimize errors and, therefore, also the central tendency.

This prediction, if verified, has important consequences that have partly been overlooked so far: if the central tendency depends on stimulus order, then experimental results using one stimulus sequence cannot be generalized to every other circumstance. Experiments demonstrating the central tendency use the paradigm introduced by Vierordt (1868): stimuli from a large range of magnitudes are randomly presented to the participant in the same context. Under natural circumstances, however, this seems to be the exception. If we have to estimate or reproduce magnitudes in daily life, subsequent stimuli in one context are similar or come from a small range. While the problem of generalizing from one experiment to the next has been recognized early on (see Hollingworth 1910; Woodrow 1930), the errors found in experiments on magnitude estimation using Vierordt’s random order protocol are quite often discussed with respect to daily life: for example, the authors of a study on facial age estimation remark that “these errors can have serious consequences” (Clifford et al. 2018).

Here we first demonstrate using the iterative updating model of magnitude reproduction (Petzschner, Glasauer, & Stephan, 2015) on Vierordt’s original data that, according to the model, the central tendency found by Vierordt vanishes when using the same stimuli but in



different order. We then confirm experimentally with a duration reproduction task that participants indeed exhibit smaller biases when stimulus order conforms to a random walk instead of being randomized.

Methods

For both parts the same model, described in Petzschner and Glasauer (2011), has been used with one difference: only one single parameter, the ratio of system to measurement noise, has been fitted to the data. Briefly, the model consists of 1) transformation of the sensory magnitudes to log space (Weber-Fechner law), 2) iterative Bayesian updating implemented as Kalman filter (Kalman, 1960) assuming that subsequent stimuli differ only by normally distributed system noise, and 3) transformation back to linear stimulus space. All software was developed using Matlab (Mathworks Inc.).

Simulation of Vierordt's results

The duration reproduction data used are taken from Table 1 (Vierordt, 1868). Since the data are reported as averages for 22 duration intervals together with the number of stimuli per interval, we iteratively constructed a stimulus set with the same properties. This set was then randomized, simulated, and used to fit the free parameter of the model to Vierordt's data. The resulting stimulus sequence was rearranged iteratively to resemble a random walk. The random walk sequence was used as model input (using the previously fitted parameter) to predict the result of a corresponding duration reproduction experiment.

Experimental validation

Participants 14 naïve volunteers (7 female, average age 27.4) participated in the experiment. The experiment was approved by the ethics committee of the Department of Psychology at LMU Munich.

Stimuli As visual stimulus a yellow disk (subtended 4.7° , 21.7 cd/m^2) was presented on a 21-inch monitor (100 Hz refresh rate) at 62 cm viewing distance using the Psychtoolbox (<http://psychtoolbox.org>).

Procedure Each trial started after 500 ms presentation of a fixation cross (Figure 1). Then the stimulus appeared for a predefined duration. After a short break of 500 ms participants were prompted to reproduce the duration of the stimulus by pressing and holding a key. At the end of the trial, a coarse visual feedback (5 categories from $<-30\%$ to $>30\%$ error) was given for 500 ms. Each participant performed two blocked sessions in balanced order. In the *random walk* session, 400 stimulus durations from 400 to 1900 ms were

presented, which were generated by a Wiener process. In the *randomized* condition, the same stimuli were used in scrambled order. Each participant received a different sequence.

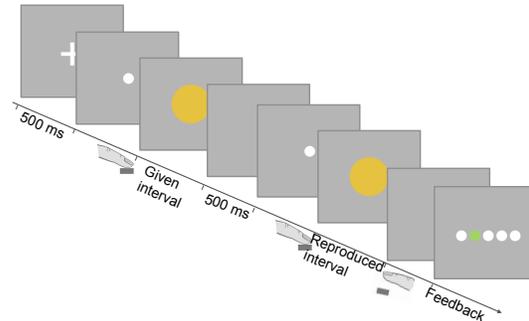


Figure 1: Experimental procedure for duration reproduction experiment.

Data analysis Reproduction was analyzed by fitting a least-squares regression to stimulus reproduction plotted over stimulus duration for each session individually to quantify the central tendency. The slope of the regression line can directly be related to the noise parameter for the *randomized* condition (Glasauer & Shi, 2018). Serial dependence was assessed by calculating the correlation between the error in trial k and the stimulus difference between trial k and $k-1$. For model simulation, the individual stimulus sequences of the *randomized* condition were used to fit the noise ratio for each participant. The fitted noise ratio was applied to predict the outcome of the *random walk* condition.

Results

Simulation of Vierordt's results

Vierordt's results and the model fit assuming a *randomized* stimulus order are shown in Figure 2 together with the prediction for the *random walk* condition. Note that stimuli used for both the fit and the prediction were exactly the same except for the order of presentation. The fitted noise ratio parameter is 6.54 and is used for both simulations.

The model prediction shows clearly that for *random walk* order of the stimulus set the bias in duration reproduction should vanish or, in other words, the central tendency would be a consequence of *randomized* stimuli.

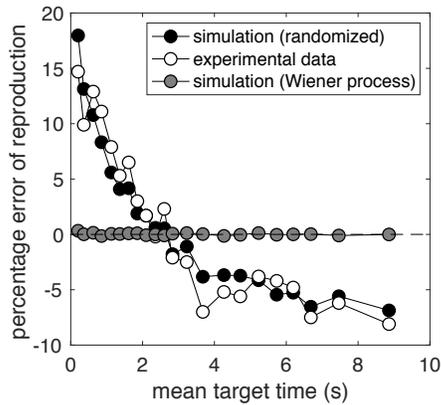


Figure 2: Vierordt's (1868) data (open circles, 1104 trials, one participant), best fit simulation of the data in *randomized* stimulus order (black), and simulation of the same stimuli in *random walk* order (gray).

Experimental results

To quantify the central tendency, we used $ci=(1-slope)$ as centrality index, with *slope* being the slope of the least-squares regression to stimulus reproduction plotted over stimulus duration. Consequently, a value of $ci=0$ indicates no central tendency and veridical reproduction. A repeated measures ANOVA of the centrality indices showed a significant effect of condition [$F(27,1)=53.5$, $p<0.0001$] with the average index close to 0 ($mean\pm SD$ 0.095 ± 0.138) for the random walk, but much higher centrality index (0.456 ± 0.173) for the randomized sequence. The average reproduction errors are shown in Figure 3 in the same format as in Figure 2 for comparison.

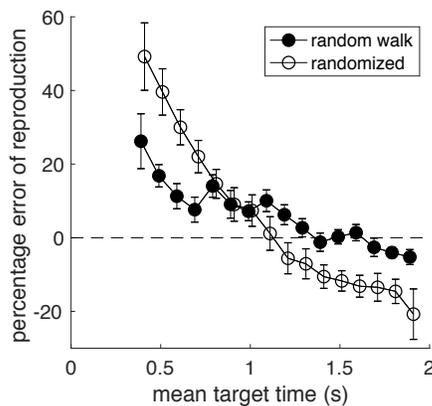


Figure 3: Average experimental results for the duration reproduction experiment ($n=14$; 400 trials per participant) for random walk (black) and randomized (white) stimulus order. Error bars denote standard error of the mean.

While our participants exhibited much higher errors than Karl Vierordt (see Figure 2), the overall result for the randomized condition is comparable. The random walk condition confirms the model prediction that errors should depend on stimulus order and decrease for the random walk. We also fitted the model individually to the randomized condition and used the parameter to predict the outcome of the random walk condition. However, the predicted errors in the random walk condition were considerably smaller than found experimentally. Closer inspection of the data showed that even though the centrality index was smaller in the random walk condition for every single subject, there were considerable differences between participants.

To assess the serial dependence, we calculated the correlation between the error in trial k and the stimulus difference between trial k and $k-1$ for the *randomized* condition. The average correlation coefficient (mean 0.53; SD 0.11) was significantly different from zero (t-test $p<0.0001$). A similar result was found for the regression slope.

Discussion

We demonstrated using the historical data provided by Vierordt (1868) that a contemporary model of magnitude reproduction (Petzschner & Glasauer 2011) predicts that the central tendency found by Vierordt and confirmed later by others (Hollingworth 1910; Lejeune & Wearden 2009; Shi et al., 2013) is a consequence of randomizing stimuli. We confirmed by a duration reproduction experiment similar to that performed by Vierordt that the central tendency indeed becomes significantly smaller for a stimulus order that imposes only small changes from one stimulus to the next and resembles a random walk.

The dependence on stimulus order and the serial dependence found also allows to distinguish two classes of models: models assuming a static prior (Jazayeri & Shadlen, 2010) vs. models assuming iterative updating (Petzschner & Glasauer 2011; Bausenhardt et al., 2014). Models with static prior would predict exactly the same results independent of stimulus order and no serial dependence, while all our participants showed a decrease of central tendency for the *random walk* condition and a strong serial dependence in the *randomized* condition.

The generative model underlying the iterative estimation process assumes that the stimulus magnitude x_k at trial k is equal to the magnitude at the previous trial plus some random amount ε , which is normally distributed with zero mean and known variance. Thus, the model is optimal, if the presented

magnitudes indeed follow a random walk. Consequently, the bias observed in randomized experiments is indeed suboptimal, because the stimulus generation does not match the model used by the perceptual estimation process (for discussion of suboptimality, see Rahnev & Denison 2018). We conclude that magnitude perception is tuned to natural circumstances, where subsequent magnitudes within the same context are similar and adhere to a random walk. Thus, the experimentally used randomized stimulus order is what is suboptimal. For the majority of participants in our experiment, the iterative (but inappropriate) model provided the best fit in the *randomized* condition. Thus, the perceptual process apparently is not flexible enough to switch to a more appropriate model, which would assume that stimuli are drawn randomly from a fixed interval, and thus would iteratively estimate that interval.

Finally, our results demonstrate that the central tendency, as suspected by others before (e.g. Woodrow 1930), is a consequence of the experimental protocol used first by Vierordt and later by many others. The results obtained in these experiments can, therefore, not be generalized to other situations, as still is customary (e.g., Clifford et al. 2018). Interestingly, Vierordt (1868) claimed that he used Fechner's "method of average error" (1860), but Fechner's descriptions of the method clearly state that the same stimulus was presented repeatedly and that, if multiple stimuli were presented in one session, the stimuli were presented in blocks with increasing or decreasing magnitude. It therefore seems that Vierordt misinterpreted Fechner's description and, by changing the experimental protocol, induced the systematic bias now known as central tendency or, in timing research, as Vierordt's law.

Acknowledgments

Supported by German DFG research projects SH 166/3-2 and GL 342/3-2. We thank Xuelian Zang and Xiuna Zhu for collecting behavioral data.

References

- Bausenhart, K. M., Dyjas, O., & Ulrich, R. (2014). Temporal reproductions are influenced by an internal reference: Explaining the Vierordt effect. *Acta Psychologica*, *147*, 60–67.
- Clifford, C. W. G., Watson, T. L., & White, D. (2018). Two sources of bias explain errors in facial age estimation. *Royal Society Open Science*, *5*, 180841.
- Fechner, G. T. (1860). *Elemente der Psychophysik. Vol I and II*. Leipzig: Breitkopf und Härtel.
- Glasauer, S., & Shi, Z. (2018). 150 years of research on Vierordt's law - Fechner's fault? *bioRxiv*, doi: 10.1101/450726.
- Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, *7*, 461-469.
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*, 1020-1026.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*, 35-45.
- Laming, D. (1999). Prior expectations in cross-modality matching. *Mathematical Social Sciences*, *38*, 343-359.
- Lejeune, H., & Wearden, J. H. (2009). Vierordt's The Experimental Study of the Time Sense (1868) and its legacy. *European Journal of Cognitive Psychology*, *21*, 941-960.
- Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *The Journal of Neuroscience*, *31*, 17220–17229.
- Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*, 285-293.
- Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*, e223:1-66.
- Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences* *17*, 556-564.
- Vierordt, K. (1868). *Der Zeitsinn nach Versuchen*. Tübingen: H. Laupp'sche Buchhandlung.
- Woodrow, H. (1930). The reproduction of temporal intervals. *Journal of Experimental Psychology*, *8*, 473–499.