PsyCh Journal
INSTITUTE OF PSYCHOLOGY
Chinese Academy of Sciences
Check for updates

PsyCh Journal **10** (2021): 732–741
DOI: 10.1002/pchj.464

The origin of Vierordt's law: The experimental protocol matters

Stefan Glasauer [1] and Zhuanghua Shi [2]

[1]Computational Neuroscience, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany, [2]General and Experimental Psychology, Ludwig-Maximilians-Universität Munich, Munich, Germany

Abstract: In 1868, Karl Vierordt discovered one type of errors in time perception—an overestimation of short duration and underestimation of long durations, known as Vierordt's law. Here we reviewed the original study in its historical context and asked whether Vierordt's law is a result of an unnatural experimental randomization protocol. Using iterative Bayesian updating, we simulated the original results with high accuracy. Importantly, the model also predicted that a slowly changing random-walk sequence produces less central tendency than a random sequence with the same durations. This was validated by a duration reproduction experiment from two sequences (random and random walk) with the same sampled distribution. The results showed that trial-wise variation influenced the magnitude of Vierordt's law. We concluded that Vierordt's law is caused by an unnatural yet widely used experimental protocol.

Keywords: Bayesian updating; central tendency; time perception, history of psychophysics; Vierordt's law
Correspondence Dr. Stefan Glasauer, Computational Neuroscience, Brandenburg University of Technology Cottbus-Senftenberg, Universitätsplatz 1, 01968 Senftenberg, Germany. Email: stefan.glasauer@b-tu.de
Received 7 October 2020. Accepted 19 April 2021.

In 1868, Karl Vierordt, professor of physiology at the University of Tübingen, published his book, *Der Zeitsinn nach Versuchen* [The Sense of Time According to Experiments] (Vierordt, 1868), just a few years after Gustav Theodor Fechner's groundbreaking book *Elemente der Psychophysik* [Elements of Psychophysics] (Fechner, 1860) and one year after Hermann von Helmholtz's *Handbuch der physiologischen Optik* [Treatise on Physiological Optics] (von Helmholtz, 1867). Vierordt's seminal book was the first quantitative attempt to investigate time perception with methodologies proposed and invented by researchers such as Ernst Weber, Gustav Theodor Fechner, and others.

One of his main findings, and the one that has survived best over time, is now known as Vierordt's law. According to this law, short temporal durations tend to be overestimated, whereas long durations tend to be underestimated. Somewhere in between there is an "indifference point" at which perceived time is veridical. The mechanisms underlying Vierordt's law have long remained obscure. Up until a decade ago, Vierordt's law was considered as an unexplained problem that "currently defies any coherent theoretical treatment" (Lejeune & Wearden, 2009).

For his main experiments—most of them were done by Vierordt himself as an only participant—his assistant produced a time interval with two clicks, and Vierordt replicated this interval by clicking a third time so that the interval between the second and third clicks was perceived as having the same duration as the interval between the first and second clicks. The machinery used for the experiments was sophisticated enough to allow recording of stimulus and response from durations of less than 250 ms to several seconds. Fortunately, Vierordt explained his methods in detail and also published most of his data summarized as tables. In the following, we concentrate on his Table A as an example (Vierordt, 1868, p. 36; see also Table 1). It lists the average stimulus duration together with the signed error of reproduction for 22 intervals (from less than 250 ms to

Table 1
Data from the original Table A, Vierordt (1868, p. 36)

Range (s)	Mean duration (s)	Reproduction error in %	Number of trials (total: 1,104)
Less than 0.25	0.204	14.7	25
0.25–0.50	0.364	9.9	49
0.5–0.75	0.626	12.9	74
0.75–1	0.856	11.1	60
1–1.25	1.129	7.9	47
1.25–1.5	1.365	5.3	54
1.5–1.75	1.614	6.5	44
1.75–2	1.854	3.0	42
2–2.25	2.099	1.7	50
2.25–2.5	2.356	−0.2	44
2.5–2.75	2.602	2.3	41
2.75–3	2.832	−2.1	35
3–3.5	3.230	−2.5	48
3.5–4	3.677	−7.0	30
4–4.5	4.264	−5.2	50
4.5–5	4.721	−5.6	89
5–5.5	5.230	−3.8	51
5.5–6	5.733	−4.2	49
6–6.5	6.194	−4.8	73
6.7–7	6.685	−7.5	44
7–8	7.462	−6.2	62
More than 8	8.860	−8.1	43

more than 8 s) and the corresponding number of repetitions (ranging from 25 to 89). Overall, the experiment included 1,104 trials presented consecutively. The results clearly demonstrated the main feature of Vierordt's law—an overestimation of the short intervals and underestimation of the long intervals with the indifference point around 2.25 s (see Figure 1A).

The method used by Vierordt (1868, p. 22) was the "method of average error" that Fechner invented (Fechner, 1860 Vol. 1, p. 120ff, Vol. 2, p. 148ff and p. 343ff) and which Müller (1904) later referred to as the "method of reproduction", now known as the method of adjustment. The method works as follows: a stimulus magnitude *N* ("*Normalreiz*," normal stimulus) is presented and followed by a test stimulus *F* ("*Fehlreiz*," error stimulus), which is adjusted by the participant, so that *N* and *F* are perceived as equal. Then the stimulus *N* is given again, followed by the adjustable *F*, and so on (Fechner, 1860, p. 190). In Volume 2, Fechner explained his method of average error in more detail (Fechner, 1860, p. 343). He applied 10 measurements of exactly the same condition (and same magnitude) consecutively. If there were multiple magnitudes, the magnitudes were tested in either increasing or decreasing order, and each magnitude was tested in a chunk of 10 measurements. Fechner was very accurate about his method: for example, concerning measurements done by a colleague, he argued that not much could be concluded from too few measurements per stimulus, which deviated from his method of average error (Fechner, 1860, p. 209).

A closer inspection of Vierordt's experiments shows various differences to the method proposed by Fechner. First of all, while Fechner was mainly interested in the just noticeable differences (JND), Vierordt reported extensively on the "constant errors," which Fechner mentioned but treated more as a side note. Other diverging aspects of Vierordt's method, already being criticized by Müller (1904), were the missing temporal exchange between *N* and *F*, and the unidirectional change of the test duration (necessarily always starting from small values). At least Vierordt partly knew that his method deviated from the one Fechner had proposed, but he defended those differences by claiming several advantages (e.g., p. 29ff and p. 35 Vierordt, 1868). However, what is easily overlooked is that according to Vierordt in the experiments, "the assistant provided … a time interval of arbitrary magnitude" (Vierordt, 1868, p. 35). According to Fechner's and Müller's descriptions, the method requires equal or ordered, rather than arbitrary, magnitudes. Thus, evidently, the method used by Vierordt was not at all what Fechner had had in mind.

Other researchers in the late 19th century confirmed Vierordt's findings (see James, 1890 Chapter XV for a

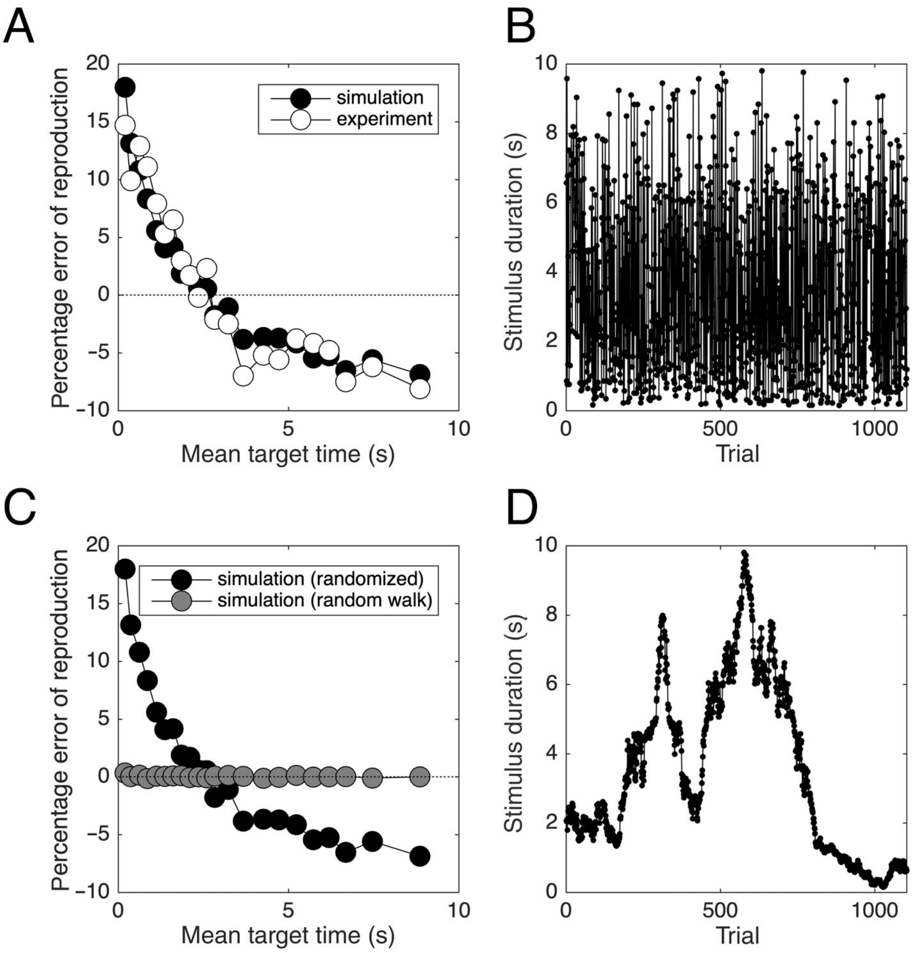

Figure 1. Reproduction data of Vierordt's durations and iterative Bayesian models. (A) data from Vierordt's original experiment (open circles) and the best fitting model simulation (filled circles). (B) The sequence of the durations used for the simulation in (A). (C) Comparison of simulation in (A) (filled circles) with the simulation (gray-filled circles) from the same sample durations but in a sequence conforming to a random walk. (D) The random-walk sequence of the same sampled durations shown in (B). (B) and (D) only differ in the sequential order.

summary), even though Fechner had already criticized their results, but mostly with respect to Weber's law (Fechner, 1884). About 40 years later, Woodrow (1930) aimed to replicate Vierordt's results but found no evidence for consistent over- and under-estimation in reproduced durations. Inspection of his methods shows that only one single interval was tested per day (50 repetitions). He explicitly mentioned: "Entirely different results might be expected from an experiment in which the various intervals were all employed on one day, particularly if they were used in an irregular order" (Woodrow, 1930, p. 476). Thus, presenting the stimuli one by one and with sufficient temporal separation, as suggested by Fechner, apparently avoids the systematic errors that are the characteristic of Vierordt's law. In other words, Vierordt's law seems to be a consequence of the particular experimental protocol.

Over the next 80 years, various other investigations followed, but without providing a formal theory for Vierordt's law. In other fields of psychophysics and experimental psychology, effects analogous to Vierordt's law were discovered for other types of magnitude estimation, such as "the law of central tendency" (Hollingworth, 1910), the "regression effect" (Stevens & Greenbaum, 1966), the "range effect" (Teghtsoonian & Teghtsoonian, 1978), and were also related to sequential or serial dependence (Cross, 1973; Fischer & Whitney, 2014; Narain, Mamassian, van Beers, Smeets, & Brenner, 2013). Interestingly, Hollingworth, who also referred to Vierordt's work, had already provided important cornerstones of the effect, such as the indifference point depending on the range of stimuli given: "in all estimates of stimuli belonging to a given range or group we tend to form our judgments around the median value of the series" (Hollingworth, 1910, p. 462). He concluded these remarkable insights from a series of experiments that he published in 1909, where he compared magnitude reproduction for different ranges of stimuli and for single stimuli presented in isolation (Hollingworth, 1909). Hollingworth's findings thus show the importance of the context of other stimuli in which a particular test stimulus is judged.

Thus, even though the basic ideas employed later in a formal theory of Vierordt's law and the central tendency (for reviews see Petzschner, Glasauer, & Stephan, 2015; Shi & Burr, 2016; Shi, Church, & Meck, 2013) were laid out early on (Hollingworth, 1910; Woodrow, 1930), explanations for these and other related phenomena, such as the range effect, have been scarce. In experimental psychology and related fields, most studies recognize and accept those types of systematic errors as a trivial finding. To our knowledge, the first study offering a quantitative formal theoretical treatment of the central tendency based on prior expectations was published at the end of the 20th century but has been completely overlooked by the scientific community (Laming, 1999). Interest, however, has been revived by the Bayesian approach for perception (e.g., Ernst & Bülthoff, 2004). We independently proposed a theory of central tendency and range effects for magnitude estimation based on iterative Bayesian inference combined with the Weber–Fechner law (Glasauer, Jahn, Stein, & Brandt, 2009; for reviews see Petzschner & Glasauer, 2011; Petzschner et al., 2015), which offered a concise explanation for the central tendency, range and order effects, and quantitatively showed how prior information on stimulus range was updated during the course of the experiment. In parallel, Jazayeri & Shadlen (2010) also proposed that the central tendency is an outcome based on integrating current sensory input with prior information about the range of the stimuli. Several similar modeling efforts followed (Acerbi, Wolpert, & Vijayakumar, 2012; Bausenhart, Dyjas, & Ulrich, 2014; Cicchini, Arrighi, Cecchetti, Giusti, & Burr, 2012; Dyjas, Bausenhart, & Ulrich, 2012; Roach, McGraw, Whitaker, & Heron, 2017). For example, Bausenhart et al. (2014) linked Vierordt's law to sequential effects and argued that it is the trial-by-trial updating process of an internal reference that leads to the central tendency. But no one took a closer look at the original data.

We hypothesized that if (1) Vierordt's law is a consequence of the experimental randomization, and (2) iterative Bayesian estimation can explain the central tendency, then we should be able to predict Vierordt's original data using our iterative Bayesian updating model (Petzschner & Glasauer, 2011) by applying the original experimental protocol as closely as possible. Randomization is now the golden method of modern psychological studies, given that repeatedly presenting the same stimulus or using a strictly ascending/descending series of stimuli, originally proposed by Fechner in his "method of average error," may introduce other types of errors, such as habituation and expectation errors. Thus, the question arises whether we have to abandon randomized stimulus presentation at all or whether we have to tolerate the range-dependent systematic errors. The answer is a direct consequence of the iterative Bayesian model (Petzschner & Glasauer, 2011), which has an important underlying assumption, namely that the magnitude/intensity of a stimulus out there in the world mostly remains constant or varies in a small range. For example, the speed of a car changes in a continuous way, not randomly jumping from one speed to another. The iterative updating model assumes that the change of the magnitude follows a Wiener process (random walk) from one trial to the next. Random walk is a common phenomenon in our daily environment, such as random motion of particles suspended in a medium, or a search path of a foraging animal. Even our eyes drift during fixation like a random walk (Engbert, Mergenthaler, Sinn, & Pikovsky, 2011). In this aspect, the model provides an optimal estimator for the slowly changing stimulus magnitude. Hence, if the sequence of the stimuli mimics a random walk, the central tendency should be greatly reduced as compared to that of a randomized sequence. On this ground, we simulated Vierordt's original study using the iterative updating model and validated the influence of experimental protocols by comparing duration reproductions under two experimental protocols: randomization and random walk.

Method

Iterative Bayesian model

The iterative Bayesian model (Glasauer, 2019; Petzschner & Glasauer, 2011) assumes that the current sensory input of the stimulus magnitude, represented by a likelihood function, is fused with a priori knowledge that is updated trial-to-trial to yield a posterior distribution, from which a most likely value is taken (e.g., maximum a posteriori probability [MAP]) as the response magnitude. To incorporate the observed Weber scaling (i.e., the standard deviation of the magnitude estimation is proportional to the absolute magnitude), we assume that the fusion and sequential updating take place at the internal log-scaled representation. After the logarithmic transformation, both likelihood and probability distributions are Gaussian (see similar approach in Ren, Allenmark, Müller, & Shi, 2020, 2021). In the reproduction stage, the posterior distribution is then transformed back into linear space to yield a response. Sequential updating is implemented via a discrete Kalman filter so that the present updating depends only on the current sensory input and the previous trial. It is assumed that the stimulus changes by a random amount from trial to trial, which is equivalent to a random walk.

The model equations (Petzschner & Glasauer, 2011) are briefly summarized here:

1. For a given trial with a stimulus magnitude of d_m, the internal representation in log-scale x_m can be expressed as $x_m = ln(d_m/d_0) + n_m$, where d_0 is an arbitrary constant and the noise in measurement $n_m \sim N\left(0, \sigma_m^2\right)$.
2. Because sensory input is noisy, magnitude estimates can be improved by integrating the prior probability of encountering a particular magnitude $N\left(x_{prior}, \sigma_{prior}^2\right)$. According to Bayesian inference (Laming, 1999), the optimal internal estimate $\hat{x}_r$ can be expressed in a linear combination of the prior and sensory measurement:

$$\hat{x}_r = w_{prior} \cdot x_{prior} + w_m \cdot x_m$$

where the weights are inverse proportional to their variance $w_{prior} = \frac{1/\sigma_{prior}^2}{1/\sigma_m^2 + 1/\sigma_{prior}^2}, w_m = 1 - w_{prior}$.

3. After the trial, we assume observers update their prior distribution, which can be captured by the Kalman filter process: the mean of the prior at trial i is: $x_{prior,i} = (1 - k_i) \cdot x_{prior,i-1} + k_i \cdot x_m$ with the Kalman gain $k_i = \frac{\sigma_{prior,i-1}^2 + \sigma_{sys}^2}{\sigma_{prior,i-1}^2 + \sigma_{sys}^2 + \sigma_m^2}$ and variance $\sigma_{prior,i}^2 = k_i \cdot \sigma_m^2$, where σ_{sys}^2 is the variance of the system noise.
4. For the reproduction, the internal estimate $\hat{x}_r$ is transformed back to the linear scale (a lognormal distribution): $d_r = e^{\hat{x}_r + \Delta x} \cdot d_0$, where Δx is a free parameter depending on which value observers select from the posterior. $\Delta x = 0$ indicates the median of the posterior, while $\Delta x = \sigma_r^2/2$ represents the mean of the posterior. For simplicity, we set $d_0 = 1$ and $\Delta x = 0$. Varying these parameters did not impact the overall estimation.

Thus, the free parameter of the model is the ratio $\sigma_{sys}^2/\sigma_m^2$, which can be alternatively approximated from the slope of the least-squares regression between stimulus and reproduction. Let d_{prior} be $e^{x_{prior}}$, it then follows with (4) and (2) that $d_r = e^{\hat{x}_r} = d_{prior}^{w_{prior}} \cdot d_m^{w_m}$. And applying the first order linearization, we get $d_r = d_{prior}^{w_{prior}} \cdot d_m^{w_m} \approx w_{prior} \cdot d_{prior} + w_m \cdot d_m$. If the weights and d_{prior} are constant, then from the linear regression $d_r = a + b \cdot d_m$ it follows that the slope is $b = w_m$ and the regression index $1 - b = w_{prior}$.

From the steady state of (3), we get $w_m = \frac{2k}{1+k}$ with k being the steady state Kalman gain and after some calculations $\frac{\sigma_{sys}^2}{\sigma_m^2} = \frac{w_m^2}{2(1-w_m)(2-w_m)}$. The steady state for the Kalman gain (and thus for the weights) is usually reached after very few trials (Petzschner & Glasauer, 2011). Note, however, that the correspondence between the regression index 1-*b* and the prior weight w_{prior} only holds if d_{prior} can be regarded as a constant. This is the case only for the randomized condition and a sufficient number of trials, in which case we can replace d_{prior} by its temporal average, which, for stimuli d_m drawn randomly from a uniform distribution will yield $d_{prior} = \mathrm{E}\{\ln(d_m)\}$. For the random walk condition, d_{prior} can no longer be regarded as constant. Thus, the regression index obtained in the random walk condition does not correspond to the prior weight.

All model simulations were programmed and performed in MATLAB (MathWorks Inc., Natick, MA, USA).

Simulation of Vierordt's data

Table 1 shows Vierordt's original study (Table A, Vierordt, 1868, p. 36). Since Vierordt binned all test durations into 22 intervals and only reported the number of trials, the range and mean of each bin and their corresponding mean reproduction errors, but not the exact durations and their sequence, we chose to iteratively find a stimulus sequence that had the same histogram as the one described by Vierordt. To do so, we randomly drew the same number of durations for each bin given by Vierordt from a uniform distribution and repeated this process until the mean difference of the sampled durations and the original report was smaller than 1 ms. This resulted in a sequence S of $\Sigma n_i = 1{,}104$ durations, which in its summary statistics accurately resembles Vierordt's stimuli. These stimuli can now be used as input for the model. However, since it is an iterative model, the exact sequence of the stimuli matters, a detail not mentioned in Vierordt's book. We therefore generated 10,000 random permutations of our 1,104 stimuli. The resulting stimulus sequences were used to fit the model (one free parameter, the system-to-measurement noise ratio $\sigma_{sys}^2/\sigma_m^2$ of the Kalman filter, see above) to Vierordt's data. For the fitting, we minimized the least-squares distance between the average result per bin obtained from the model simulation and the average result per bin reported by Vierordt. Thus, for each randomized stimulus sequence, we obtained one parameter reflecting the best fitting model. We then selected the stimulus sequence S_{min} that resulted in the model simulation with the smallest least squares error with respect to the reported results for depiction in Figure 1A,B. Note that the only difference between S and S_{min} is the order of stimuli.

The same sequence S was then reshuffled to yield a new sequence that mimics a random walk. To do so, another iterative procedure was used. First, a random-walk sequence R of 1,104 values was created by integrating normally distributed random numbers. The resulting random walk R was normalized to the same range as the original stimulus sequence S in Figure 1B. Both sequences were then sorted and the mean least-squares distance between sorted sequences R_{sort} and S_{sort} was calculated. A perfect match would thus mean that S, when properly sorted, would exactly resemble the true random walk R. This procedure was repeated 1,000,000 times and the sequence R for which S_{sort} best resembled R_{sort} was selected. The sorted sequence S_{sort} was then brought into the appropriate order by un-sorting it according to the sort order of R_{sort}. Note again that S_{sort} contains exactly the same values as S. The properly sorted new sequence S_{rw}, which is closely resembling a random walk, is depicted in Figure 1D and used for model simulation in Figure 1C (same model parameter as in Figure 1A).

Validation experiment

In order to validate the influence of experimental protocols (e.g., random vs. random walk) on the central tendency effect, we conducted a duration reproduction task. In the experiment, participants received two sessions that had the same probe durations, but differed in their sequences. One sequence was in random order, while the other sequence had a random-walk structure. The random session has the same randomization protocol as Vieordt's original study, though in a short range of durations, while the random-walk condition aimed to validate the iterative updating model.

Participants

Fourteen volunteers (seven female, average age 27.4 years) participated in the experiment. All participants had normal or corrected-to-normal vision. They were naive to the purpose of the experiment and were monetarily compensated for their participation (9 Euro/h). The experiment was approved by the ethics committee of the Department of Psychology at Ludwig-Maximilian University Munich. Informed consent was obtained prior to the experiment.

Apparatus and stimuli

The experiment was conducted in a sound-isolated, dimly lit cabin (5 cd/m^2). The visual stimulus was a yellow disk (subtended 4.7°, 21.7 cd/m^2), presented on a 21-in. LACIE CRT monitor, with a refresh rate of 100 Hz. The viewing distance was about 62 cm by asking participants to sit at a predefined fixed position. The experimental program was developed using MATLAB (MathWorks Inc.) and Psychtoolbox (http://psychtoolbox.org).

Design and procedure

We adopted a duration production-reproduction task, which has been commonly used in previous studies (e.g., Lewis & Miall, 2009; Ren et al., 2021; Shi, Ganzenmüller, & Müller, 2013). Each trial started with the presentation of a center fixation cross for 500 ms, which was then replaced by a white dot. The white dot prompted participants to press the down arrow key to initiate the production phase. Immediately after the key press, a yellow disk was presented in the center of the screen for a given sampled duration (ranging from 400 ms to 1900 ms). Participants were instructed to hold the response key as long as the stimulus was on and release the key when the stimulus was off. This was to maintain the participant's attention to the presented duration. The blank screen remained for 500 ms after the key release, which was followed by a second white dot at the center of the screen, prompting the reproduction phase. Participants were asked to press and hold the response key as long as the perceived duration that was shown in the production phase. A feedback display was presented for 500 ms at the end of each trial, showing participants' relative reproduced error (i.e., ratio of the reproduced error to the given duration) by highlighting one of five horizontally arranged circles. The five circles from left to right mapped to the relative error ranges: less than –30%, from −30% to −5%, from −5% to 5%, from 5% to 30%, and more than

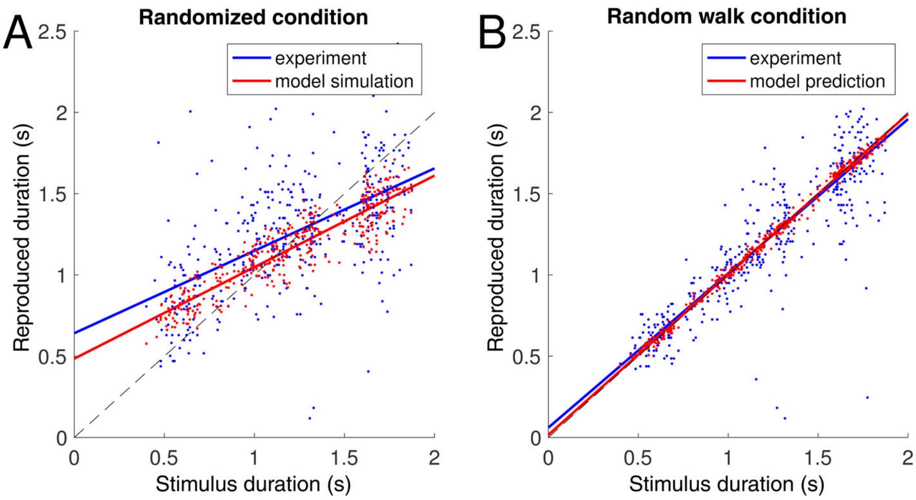

Figure 2. Representative data from one subject together with model simulations. (A) experimental data from the randomized condition (blue dots) together with linear regression line (slope 0.507) and model simulation (best fit, red dots) with regression line (slope 0.563). Note that the offset between regression lines could be avoided by fitting the location parameter of the model. (B) Experimental data from the random walk condition (blue dots) together with linear regression line (slope 0.949) and model prediction (using the fitted parameter from (A), red dots) with regression line (slope 0.987).

30%, respectively. The middle three circles were shown in green, and the right and left circles were shown in red, indicating that a large reproduction error was made.

Each experiment consisted of two sessions, with one session using the random-walk sequence and the other session using the randomized sequence. A random-walk sequence of 400 durations was first generated for each participant using a Wiener process in the range of 400 to 1900 ms. The randomized session used the same durations but in a randomized order. The order of the two sessions was counterbalanced among participants.

Data analysis

Stimulus durations were categorized into stimulus intervals of 100 ms. The average percent error of reproduced duration in each interval for both conditions were calulated (see Results, Figure 3A). As a measure of the central tendency, we used the slope of the linear regression line fitted to the reproduced duration over the stimulus duration. A unity slope would thus indicate independently of constant over- or underestimation that no central tendency is present. As outlined above, the slope of the regression line can be directly interpreted in terms of the Bayesian model for the randomized order.

Results and discussion

Model simulation

Figure 1A depicts Vierordt's original data together with the best fit from the simulation, and Figure 1B shows the best simulated sequence. Evidently, the model provides an excellent fit to Vierordt's data. However, how much does the reproduction error depend on experimental protocols? Assuming the same intervals are provided in ascending or descending order (assuming the same model with identical parameters), the model predicts that the absolute percentage error would be below 0.2% for all intervals (as compared to below 15% in Vierordt's original data). The differential outputs of the simulation corroborate our suspicion that Vierordt's law is a consequence of the random presentation of stimuli within the same experimental context. The iterative Bayesian updating model thus can explain both Hollingworth's conclusions about the central tendency and Woodrow's failure replication of Vierordt's findings.

Figure 1C illustrates this difference between a random-walk sequence (Figure 1D) and a randomized sequence (note that stimuli in 1B and 1D are the same except for the temporal order of presentation). As we predicted, the central tendency was almost suppressed with the random-walk sequence.

Experimental validation

Figure 2 depicts one representative participant for two experimental protocols. By visual inspection, the randomization protocol yielded a strong central tendency effect (linear slope: 0.507), while the random-walk sequence produced little systematic bias (slope: 0.949). We also applied the dynamic-updating Bayesian model to the reproduction data from the "randomized" condition for each participant separately and used the fitted model parameter to predict the results from the "random walk" (the red lines in Figure 2). The iterative Bayesian updating model provided good predictions for both conditions.

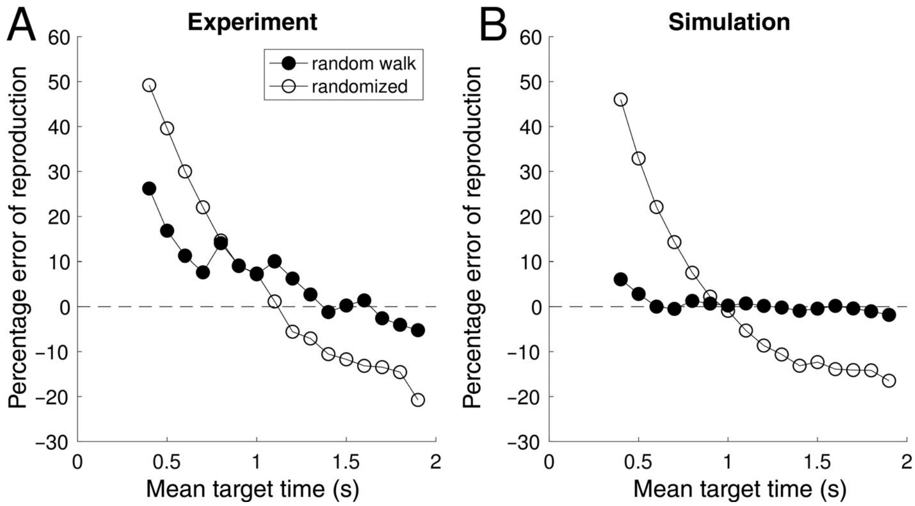

Figure 3. Reproduction experiment with stimuli ranging from 0.4 s to 1.9 s (400 durations per subject per condition). (A) Average reproduction error as a function of the duration from 14 subjects in two experimental conditions (open circles: randomized; closed circles: random walk). (B) The simulation of the iterative Bayesian model for the "randomized" data (open circles, best fit for each participant separately) and prediction for "random walk" data (closed circles, using the individual model parameters from the randomized experiment). Error bars show standard error of the mean.

The mean percentages of reproduction errors are shown in Figure 3A, which clearly indicates the central tendency (i.e., Vierordt's law). While Vierordt's law is still visible for the random walk sequence, the average errors are drastically lower than those in the randomized sequence, despite both conditions containing the same set of stimuli, but presented in a different order. We fitted a linear regression for each condition in each individual dataset, using the fitted slope as the regression index. A repeated-measures analysis of variance of the regression indices revealed a significant effect of randomization [$F(1, 13) = 46.48$, $p < .0001$, $\eta_p^2 = 0.781$] with the average index close to 0 (mean $\pm$ SD: 0.092 ± 0.138) for the random walk, but much higher regression index (0.447 ± 0.167) for the randomized sequence.

The predicted results from the dynamic-updating Bayesian model are shown in Figure 3B. A comparison of actual and predicted data shows that the average central tendency in the random-walk condition is higher than predicted from the model simulation (predicted average regression index 0.011 ± 0.011). Closer inspection shows that the learning rate for the prior in the random-walk condition is, at least for some participants, slower than expected from the model fitted to the randomized condition. However, overall, there is a good correspondence between data and prediction: the central tendency does not vanish but becomes significantly smaller.

General discussion

Here, we traced Vierordt's law back to his original study and its experimental protocol, which deviates from Fechner's proposal. Reviewing debates on Vierordt's law over one and a half centuries, we hypothesized that the major factor for Vierordt's law is the experimental protocol. We then applied an iterative updating Bayesian model to simulate the original results and conducted an experiment to compare two experimental protocols: randomization and random walk. Both simulation and behavioral data confirmed our hypothesis—the experimental protocol (here, sequence) matters.

While other factors still play a role—especially the range of stimuli presented—systematic errors can thus be minimized by an appropriate stimulus design even without abandoning randomization completely, but instead resorting to random walks. Notably, our experiment also extends Hollingworth's claims (Hollingworth, 1909, 1910): it is not just the range of magnitudes presented that determines the central tendency, but even more so their sequence. Moreover, our findings refute models that assume a *static* prior distribution depending on the range of stimuli as a reason for the central tendency (e.g., Cicchini et al., 2012; Jazayeri & Shadlen, 2010). According to the *static* models, the same results are expected for both sequences, given that only the range and the sample distribution matter, while the sequential order plays no role in it. By contrast, the iterative updating model assumes the prior is updated from trial to trial, thus the central tendency depends on the structure of the sequence—more marked central tendency for the random sequence as compared to the random-walk sequence.

The random-walk sequences follow a Wiener process, which has commonly been found in our environment, for example, drift diffusion of physical particles, fixational drifts of our eyes (Engbert et al., 2011), and information accumulation for decision-making (Gold & Shadlen, 2007; Smith & Ratcliff, 2004). Random walks reflect the nature of small fluctuations of a dynamic stochastic process—our world is relatively stable but fluctuates in small changes. The iterative updating model, indeed, incorporates this assumption—it generates a short-term expectation of the next stimulus based on the recent exposed context in the prior distribution. In this aspect, the reduction of the central tendency effect is a result of correct expectation from the internal prediction based on the previous stimuli. The random fluctuation in the random walk matches the continuous updating, which also reduces the central tendency bias. In contrast, habituating a prior or, in other words, learning the complete stimulus distribution, would increase the central tendency for random-walk stimuli.

It is interesting to note that the central tendency bias predicted by the iterative model is smaller than what we observed in the validation study (see Figure 3). Subject-wise analysis revealed that for some participants, the updating rate of the prior in the random-walk session is slower than the model predicted based on the randomized session. It suggests observers might consider additional factors, such as the range and the distribution of the stimuli, other than the sequence itself in updating the priors. Thus, the observed bias is somewhat between the predictions of the pure iterative updating model (Glasauer, 2019; Petzschner & Glasauer, 2011) and the static models (e.g., Cicchini et al., 2012; Jazayeri & Shadlen, 2010).

In summary, from a re-evaluation of the original dataset with iterative Bayesian modeling and validation by new experiments we conclude that Vierordt's law (and the central tendency) is a result of the specific experimental protocol—randomly presenting stimuli with large trial-to-trial magnitude fluctuation. This protocol deviates from what usually happens in everyday life, where either successive magnitudes are equal and share the same context, or different magnitudes are associated with different contexts. The proposed underlying mechanism of Bayesian dynamic updating indeed improves performance over trials for equal or slowly changing magnitudes but not for rapid large magnitude fluctuations. According to our analysis, 150 years of research on Vierordt's law have thus focused on an effect that is caused by an unnatural but since then widely adopted experimental protocol, which was first introduced by Vierordt, who misinterpreted the method of reproduction invented by Fechner and described in his groundbreaking *Elemente der Psychophysik* (Fechner, 1860).

Disclosure of conflict of interest

The authors declare there are no conflicts of interest.

Acknowledgments

This study was 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

Acerbi, L., Wolpert, D. M., & Vijayakumar, S. (2012). Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Computational Biology*, *8* (11), e1002771. https://doi.org/10.1371/journal.pcbi.1002771

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. https://doi.org/10.1016/j.actpsy.2013.06.011

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*, *32*(3), 1056–1060. https://doi.org/10.1523/JNEUROSCI.3411-11.2012

Cross, D. V. (1973). Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics*, *14*(3), 547–552. https://doi.org/10.3758/BF03211196

Dyjas, O., Bausenhart, K. M., & Ulrich, R. (2012). Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. *Attention, Perception & Psychophysics*, *74*(8), 1819–1841. https://doi.org/10.3758/s13414-012-0362-4

Engbert, R., Mergenthaler, K., Sinn, P., & Pikovsky, A. An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences*. 2011; *108*(39):E765–E770. http://doi.org/10.1073/pnas.1102730108.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Fechner, G. T. (1860). *Elemente der Psychophysik*. (Vol. *I and II*). Leipzig, Germany: Breitkopf und Härtel.

Fechner, G. T. (1884). In Ueber die Frage des Weber's chen Gesetzes und Periodicitaetsgesetzes im Gebiete des Zeitsinnes. *Abhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften*, (Vol. 22). Leipzig, Germany: S. Hirzel.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Glasauer, S. (2019). Sequential Bayesian updating as a model for human perception. In S. Ramat & A. G. Shaikh (Eds.), *Mathematical Modelling in Motor Neuroscience: State of the Art and Translation to the Clinic. Gaze Orienting Mechanisms and Disease*. (Vol. 249), pp. 3–18. Amsterdam: Elsevier. https://doi.org/10.1016/bs.pbr.2019.04.025

Glasauer, S., Jahn, K., Stein, A., & Brandt, T. (2009). On the origin of systematic errors in a simple navigation task. *BMC Neuroscience*, *10*(S1), P13–P13. https://doi.org/10.1186/1471-2202-10-s1-p13

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Hollingworth, H. L. (1909). *The inaccuracy of movement: With special reference to constant errors*. (Archives of Psychology, No. 13). New York, NY: Science Press.

Hollingworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology and Scientific Methods*, *7* (17), 461–469. https://doi.org/10.2307/2012819

James, W. (1890). *The principles of psychology*. (Vol. *1*). New York, NY: Henry Holt and Co.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. https://doi.org/10.1038/nn.2590

Laming, D. (1999). Prior expectations in cross-modality matching. *Mathematical Social Sciences*, *38*(3), 343–359. https://doi.org/10.1016/S0165-4896(99)00024-4

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*(6), 941–960. https://doi.org/10.1080/09541440802453006

Lewis, P. A., & Miall, R. C. (2009). The precision of temporal judgement: Milliseconds, many minutes, and beyond. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1525), 1897–1905. https://doi.org/10.1098/rstb.2009.0020

Müller, G. (1904). *Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik*. Wiesbaden, Germany: J.F. Bergmann.

Narain, D., Mamassian, P., van Beers, R. J., Smeets, J. B. J., & Brenner, E. (2013). How the statistics of sequential presentation influence the learning of structure. *PLoS One*, *8*(4), e62276. https://doi.org/10.1371/journal.pone.0062276

Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *Journal of Neuroscience*, *31* (47), 17220–17229. https://doi.org/10.1523/JNEUROSCI.2028-11.2011

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*, 1–9. https://doi.org/10.1016/j.tics.2015.03.002

Ren, Y., Allenmark, F., Müller, H. J., & Shi, Z. (2020). Logarithmic encoding of ensemble time intervals. *Scientific Reports*, *10* (1), 18174. https://doi.org/10.1038/s41598-020-75191-6

Ren, Y., Allenmark, F., Müller, H. J., & Shi, Z. (2021). Variation in the "coefficient of variation": Rethinking the violation of the scalar property in time-duration judgments. *Acta Psychologica*, *214*, 103263. https://doi.org/10.1016/j.actpsy.2021.103263

Roach, N. W., McGraw, P. V., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences*, *114*(2), 412–417. https://doi.org/10.1073/pnas.1610706114

Shi, Z., & Burr, D. (2016). Predictive coding of multisensory timing. *Current Opinion in Behavioral Sciences*, *8*, 200–206. https://doi.org/10.1016/j.cobeha.2016.02.014

Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, *17*, 1–9. https://doi.org/10.1016/j.tics.2013.09.009

Shi, Z., Ganzenmüller, S., & Müller, H. J. (2013). Reducing bias in auditory duration reproduction by integrating the reproduced signal. *PLoS One*, *8*(4), e62065. https://doi.org/10.1371/journal.pone.0062065

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168. https://doi.org/10.1016/j.tins.2004.01.006

Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Attention, Perception & Psychophysics*, *1*(5), 439–446. https://doi.org/10.3758/BF03207424

Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, *24*(4), 305–314. https://doi.org/10.3758/BF03204247

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen*. Tübingen, Germany: Laupp.

von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik*. Hamburg, Germany: Leopold Voss.

Woodrow, H. (1930). The reproduction of temporal intervals. *Journal of Experimental Psychology*, *13*(6), 473–499.