Individual believes about temporal continuity explain perceptual biases

Stefan Glasauer[1*] & Zhuanghua Shi[2]

[1]Computational Neuroscience, Brandenburg University of Technology Cottbus-Senftenberg, 01968 Senftenberg, Germany
[2]Experimental Psychology, Department of Psychology, Ludwig-Maximilians-University Munich, 80802 Munich, Germany

ORCID
Stefan Glasauer: 0000-0002-4313-6210
Zhuanghua Shi: 0000-0003-2388-6695

Corresponding Author
Stefan Glasauer
Address: BTU Cottbus-Senftenberg
Universitätsplatz 1
01968 Senftenberg
Deutschland
Phone +49 3573 85614
E-mail: stefan.glasauer@b-tu.de

Keywords: perceptual bias, Bayesian model, time perception, distance perception, central tendency, serial dependence

Abstract

Perceptual biases vary considerably between individuals. In the framework of probabilistic perception, these variations are commonly attributed to differences in sensory noise, which determines reliance on internal priors, and thus the size of biases. However, sensory noise is not the only determinant of perceptual outcomes: internal generative models, which express our believes about how stimuli are generated in the world, play a decisive role. These believes are mirrored in the types of explanatory models, static or iterative, offered in the literature. While static models are based on the assumption that consecutive stimuli are independent, iterative models presume some temporal continuity. Here we compare experimental results for time and distance estimation with model predictions and propose that interindividual differences cannot be explained by individual levels of sensory noise alone, but that differences in biases such as central tendency and serial dependence are based on individual believes expressed by different generative models.

Introduction

Magnitude estimates are pervasive in our daily activities, such as predicting forthcoming events, estimating the travel distance, and precisely controlling our movements. Through the history of psychophysics, several perceptual biases in magnitude perception have been found. Two of them, the central tendency effect (Vierordt 1868; Hollingworth 1910) and the sequential dependence (Holland & Lockhead 1968; Cross 1973), are still hotly debated till today Petzschner et al. 2015; Shi et al. 2013). The central tendency effect refers to a systematic overestimation of small magnitudes and underestimation of large magnitudes, whereas the sequential dependence shows that the current perceptual estimate depends not just on the current stimulus, but also on stimuli given in the past. While both biases have long been accepted as inevitable properties of magnitude perception, several quantitative theoretical accounts only emerged in the last decade that linked them to Bayesian inference of perception (Jazayeri & Shadlen 2010; Petzschner & Glasauer 2011; Cicchini et al. 2012; Roach et al. 2017). The concept of Bayesian inference also provides an operational explanation for interindividual differences seen in perception: Bayesian inference considers a perceptual estimate to result from a near-optimal combination of sensory inputs and prior knowledge. The extent to which sensory input and prior knowledge are weighted depends on the magnitude of sensory noise and the certainty of prior knowledge. When driving through fog, knowledge of the road ahead is more important than on a sunny day. Thus, individual sensory variability and certainty about the prior knowledge is commonly thought to explain inter-individual differences (Petzschner & Glasauer 2011; Powell et al. 2016).

However, Bayesian perception also offers an alternative account of individual differences: the underlying internal generative model (not to be confused with the explanatory models discussed below) that expresses the assumptions of how sensory stimuli are caused, or generated, in the external world. For example, we may assume that consecutive stimuli presented to us in an experiment are drawn randomly from a fixed distribution, just like throwing dice. Under this assumption, consecutive stimuli are assumed to be independent. If the stimulus range is known,

the perceptual estimate can be improved by combining this fixed prior knowledge with the sensory measurement. This explanation was suggested, for example, by Jazayeri and Shadlen (2010) to account for the central tendency in duration reproduction. By contrast, Petzschner & Glasauer (2011) in a distance reproduction study proposed that the central tendency is a consequence of an iterative Bayesian process that updates the internal stimulus statistics from trial to trial rather than assuming a fixed prior. The underlying generative model in this case assumes that consecutive stimuli are not independent but adhere to some temporal continuity. In the following years, various other similar explanations have been proposed, some of them using iterative updating (Dyjas et al. 2012; Thurley 2016), others assuming static priors (Cicchini et al. 2012; Roach et al. 2017). The common idea of these models is that the central tendency is a by-product of optimizing the reliability of estimates using additional cues when current sensory input is corrupted by noise.

However, static and iterative models are based on different suppositions about how the sequential structure of the stimuli is generated, and thus about the underlying generative models. Investigating sequential dependence can thus not only help distinguishing between the two cases - sequential dependence is only predicted if temporal continuity is assumed – but also help resolving whether individual differences are due to different amounts of sensory or prior variability, or due to different believes about the generative process of stimuli in the world.
In the following we thus examined central tendency and serial dependence in three experimental data sets of magnitude estimation: one from duration reproduction (Glasauer & Shi 2019, 2021), and two from linear and angular distance reproduction (Petzschner & Glasauer 2011, data set published as Petzschner & Glasauer 2020).

Results

In the literature, two main methods to quantify sequential dependence can be found. Older studies such as Holland and Lockhead (1968) quantified the sequential effect as the dependence of the current error on the stimulus magnitude in the previous trial (we refer to it as the absolute sequential dependence, ASD). More recent studies (e.g., Fischer & Whitney 2014, Bliss et al. 2017, Kiyonaga et al. 2017, Clifford et al. 2018, Cicchini et al. 2018) reported the dependence of the current error on the difference between the stimuli in the previous and current trials (the relative sequential dependence, RSD).
Even though it is rarely mentioned, the RSD is appropriate only when stimuli 1) come from a circular scale, such as angular orientation, and 2) are equally distributed over the whole scale. This is the case for most of the studies mentioned above, which investigated serial dependence for perception of visual orientation. For other cases, such as when stimuli are drawn from an open scale or from only part of a circular scale, the RSD is problematic and potentially misleading, because it inflates the true sequential dependence effect. Moreover, the RSD can then falsely show a dependence, based on mathematical coupling (Archie 1981, Curran-Everett 2010), even if the result of the current trial is completely independent of the previous stimulus (see SI Appendix A1 and A2).

In the following, we thus quantified serial dependence via the ASD (current error depending on previous stimulus, Holland and Lockhead 1968). Another possible quantification would be the cross-correlation between stimulus and reproduction, which, in case of true serial dependence, should yield significant values for lag 1 or higher. However, the cross-correlation (as does the correlation) reports the strength of the relation, but not the relation itself, and is therefore less suited for quantifying serial dependence.

Central tendency and serial dependence for static and iterative models

Given a set of stimuli x_i drawn from a normal distribution on an open scale with mean $\bar{x}$, a simple *static* model for the perceptual response y_i would be:

$$y_i = wx_i + (1 - w)\bar{x} \tag{1}$$

with the weight w being determined by using the variance of the stimulus distribution and the variance of the measurement noise. Note that the model assumes that y_i only depends on the current stimulus x_i, but not on the previous one. The fixed prior of the model could be the mean of the stimulus distribution $\bar{x}$. In this model, the central tendency is given as $c = 1 - w$. Since in this model the current response does not depend on the previous stimulus, the serial dependence is zero regardless of the central tendency (see SI Appendix A2).

For an *iterative* model, the quantification of serial dependence should yield an effect, given that in such a model the actual response is defined to depend on both the current and the previous magnitudes. The simplest iterative Bayesian model can be derived from two assumptions for the underlying generative process (Glasauer 2019): 1) the stimulus at the current trial is the same as the one on the previous trial plus some random fluctuation and 2) the sensation of the stimulus is corrupted by measurement noise. For normally distributed fluctuations and noise, the Bayesian optimal estimator model can be written as Kalman filter. When the Kalman gain k of the model reaches its steady state (usually after few trials), its equations simplify to a weighted average, so that the response y_i at trial i becomes

$$y_i = kx_i + (1 - k)y_{i-1} \tag{2}$$

with x_i being the current measurement of the stimulus and y_{i-1} the estimate at the preceding trial $i - 1$ (Glasauer 2019). Note that for a fixed k this model is equivalent to the so-called "internal reference model" (Dyjas et al. 2012, Bausenhart et al. 2014).

For this iterative model, the relationship between the central tendency and the ASD can be determined analytically (see SI Appendix A3). The central tendency is given as $c = 1 - k$. Intuitively, the extreme case with $k = 0$ causes the current response to completely depend on the initial response (which may be arbitrary) and does not change at all; therefore, the ASD becomes zero. On the other extreme, with $k = 1$ the response is veridical, always equal to the current stimulus, and independent of the previous, which yields zero serial dependence. The maximum expected absolute serial dependence is 0.25 for central tendency 0.5 (see also Fig. 5 below). Thus, for central tendencies found experimentally, there should be a distinct testable difference between the static model (serial dependence 0 and independent of central tendency) and the simple iterative model.

Using sequential dependence to dissociate generative Bayesian models

Before proceeding to experimental tests, we reconsider the difference between the generative assumptions of the static and iterative models. In both models, measurement noise r corrupts the actual sensory input. Thus, it is helpful to estimate the stimulus using additional prior information.

- The *static* model assumes that the stimulus x_i in trial i comes from a distribution $D(m, v)$ with a constant mean m and variance v. We thus can write the generative model as $x_i = m + \varepsilon_x$ with ε_x being a random number coming from a distribution $D(0, v)$.
- The *simple* iterative model assumes that the stimulus x_i in trial i is the same as in trial $i - 1$ except for some random change with variance q. In other words, the generative model is $x_i = x_{i-1} + \varepsilon_m$ with ε_m coming from a distribution $D(0, q)$.

From these assumptions we can construct a third generative model, the two-state model, that combines advantages of both models (see Materials and Methods for the details of the model):

- The *two-state* model assumes that the stimulus x_i in trial i comes from a random distribution $D(m_{i-1}, v)$ with mean m_{i-1} and variance v. The mean of this distribution in trial i is the same as in trial $i - 1$ except for some random change with variance q. In other words, the stimulus distribution in the current trial depends on that in the previous trial. The generative model now has two states: the randomly changing mean of the stimulus distribution $m_i = m_{i-1} + \varepsilon_m$ and the actual stimulus $x_i = m_{i-1} + \varepsilon_x$, drawn from this distribution.

Fig. 1 schematically illustrates the trial-to-trial difference of the three models. Each model has a distinct mechanism of estimation process, which yields different predictions concerning the absolute serial dependence. The two-state model assumes two hidden states (the stimulus and the mean of the stimulus distribution, Fig. 1B), which generalizes the *two-stage* model proposed previously (Petzschner & Glasauer 2011). The original *two-stage* model assumes that the variance v of the stimulus distribution is directly related to the variance of the known random fluctuations q of the mean and that it could be estimated from these fluctuations. When this restriction is released, the resulting two-state model encompasses the *two-stage* model as a special case, and also both the static model and the simple iterative model as two boundary cases: The static model is a boundary case of the two-state model[*], if the mean m_i in the two-state model is constant instead of fluctuating (i.e., $q = 0$). When the assumed stimulus distribution has a negligible variance ($v = 0$), the two-state model reduces to the simple iterative model.

[*] To be exact, the two-state model with q=0 and the static model are equivalent only for the steady state, because the two-state model initially estimates the unknown mean of the stimulus distribution m, while the static model assumes that the mean m is already known at trial 1.

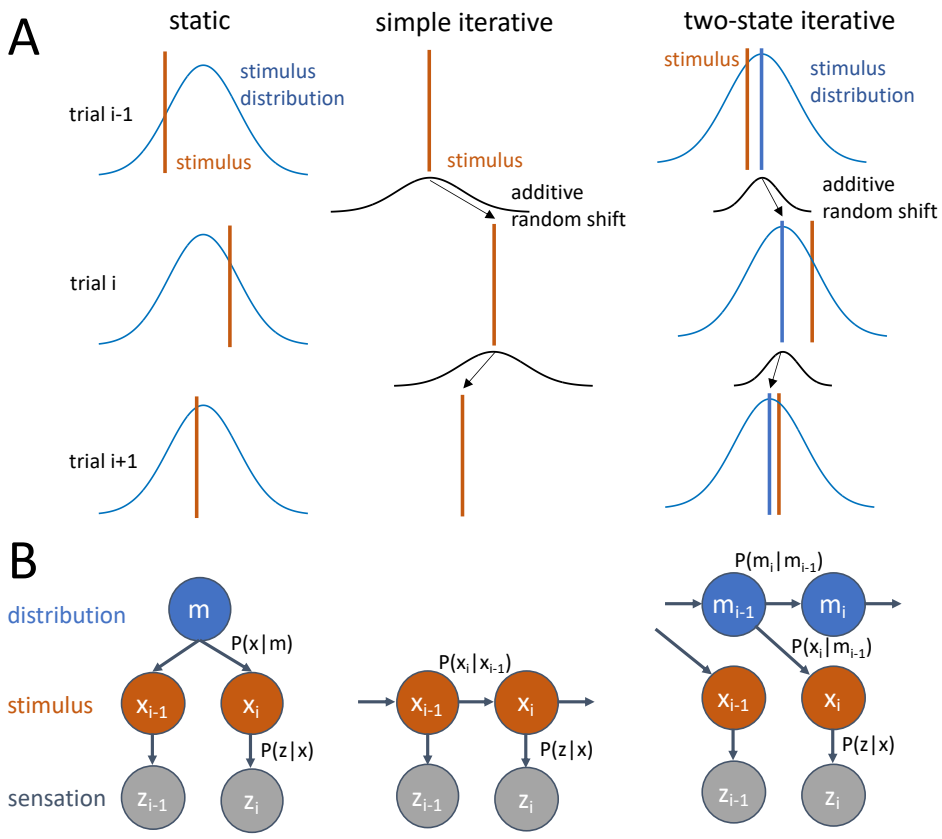

Figure 1: Generative models of magnitude reproduction. **A**: The static model (left column) assumes that stimuli (red lines) are randomly drawn from a fixed stimulus distribution (blue). In the simple iterative model (middle column), the stimulus at the current trial i is the same as in the previous trial i-1 except for some random fluctuation that is conceptualized as additive random shift (black distribution). The two-state iterative model (right column) combines these two models: the stimulus is drawn from a stimulus distribution (blue), but the mean of the stimulus distribution fluctuates from trial to trial by a random amount (additive random shift, black distribution). The stimulus itself is the same in all three cases, but the assumed sequential structure differs in each model. The measurement noise distribution around the stimuli has been omitted for clarity. **B**: Hidden Markov Model diagram visualizing the three generative models. Note that the sensation z, the observable, is also shown. It is derived from the stimulus via the conditional probability $P(z|x)$, which describes the measurement noise. For the static model, the state m, which represents the mean of the stimulus distribution, is assumed not to change from trial to trial. Stimuli are drawn from the stimulus distribution with $P(x|m)$. In the simple iterative model, stimuli change from one trial to the next with transition probability $P(x_i|x_{i-1})$. In the two-state model, the mean of the stimulus distribution changes from trial to trial with transition probability $P(m_i|m_{i-1})$, and the stimuli are drawn from the stimulus distribution with $P(x_i|m_{i-1})$.

As we have seen, the three models make different assumptions about how a specific sequence of stimuli is generated (Fig. 1). As an example, Fig 2 shows histograms, autocorrelation, and time course of exemplary stimulus sequences generated with the three models. The stimulus sequences of the iterative models have been generated so that their histograms show as much similarity as possible to the histogram of the static model's sequence (quantified by minimizing the Kullback-Leibler divergence). While the histograms are reasonably similar (Fig. 2A, KL divergence < 0.01), the autocorrelation differs considerably (Fig 2B). As expected, the sequence for the static model, which is a Gaussian noise sequence, shows no dependence between current and previous values, while the two iterative models generate sequences with autocorrelation at higher lags. The sequence generated by the simple iterative model is a Wiener process or

random walk, while the sequence of the two-state model is a superposition of a random walk and Gaussian noise. The corresponding exemplary time courses are shown in Fig. 2C, in which the blue represents the time course generated by the static model, and the red random walk time course correspond to the simple iterative model, and the yellow trace, generated by the two-state model, is a compromise between the randomness of the static model and the slow drift of the simple iterative model. The generative models thus assume quite different stimulus time courses.

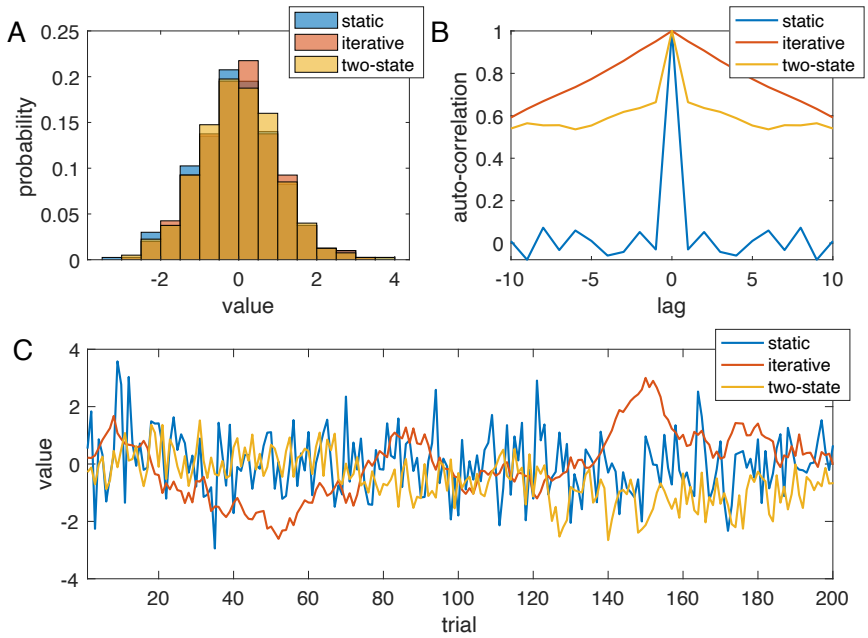

Figure 2: Sequences of stimuli generated by the three models. The sequences for the iterative and two-state model have been generated such that their histograms resemble that of the static sequence. A: histograms. B: autocorrelation. C: first 200 samples of all three sequences. The variances of the two-state model were chosen so that the autocorrelation at lag 1 equals approx. 0.66.

Experimental validation of serial dependence and the two-state model

The first dataset comes from a duration reproduction study (preliminary data were presented in Glasauer & Shi 2019, 2021). In the experiment, subjects (n=14, 7 female, average age 27.4) had to reproduce a visually presented stimulus duration by pressing and holding a key. Each subject received a random sequence of 400 stimulus durations from 400 to 1900 ms. We quantified both the ASD and the central tendency effect using simple linear regression (see Materials and Methods for the detailed method). Fig. 3 show an example of individual raw results plotted for evaluation of central tendency (Fig. 3A) and serial dependence (Fig. 3B). The relation of ASD and central tendency for all individual participants is depicted in Fig. 4A together with the predicted relation for the simple iterative model (red curve). Individual responses show a large scatter both for central tendency and serial dependence but lie all within the possible range for the two-state model. The mean serial dependence was 0.108±0.056 (mean ± SD) and significantly different from zero ($p<.0001$; t-test, n=14), which ruled out the static model for this experiment. In fact, all data points show higher serial dependence than predicted by the static model (close to zero). We conducted a partial correlation analysis and calculated

the correlation coefficient between current error and previous stimulus after controlling for the current stimulus. The average partial correlation coefficient was 0.197±0.068 and significantly different from zero (p<.0001; t-test, n=14). For comparison, the corresponding partial correlation coefficient between error and current stimulus, after controlling for the previous stimulus, was -0.623±0.156 (p<.0001; t-test, n=14), revealing the central tendency.

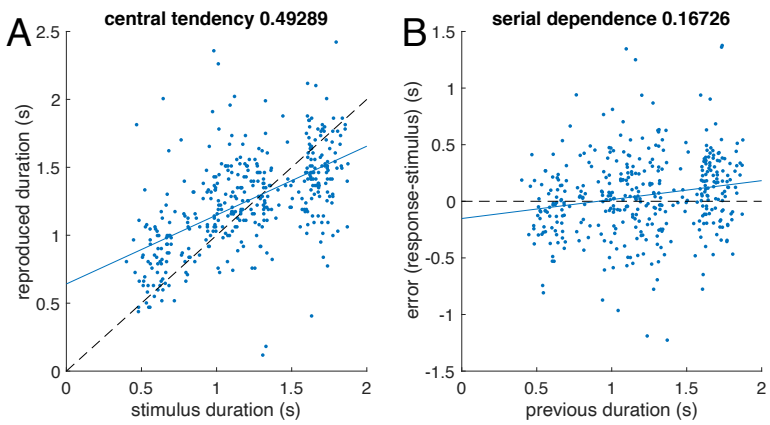

Figure 3: Data (blue dots) from one subject (400 trials) in a duration reproduction experiment. Stimuli were presented in random order. **A**: Reproduced duration plotted over stimulus duration; the central tendency is shown by the regression line (blue). **B**: Error (stimulus-response) in the current trial plotted over the duration of the previous trial. The absolute serial dependence is shown by the blue line.

Individual ASDs not only show values higher than zero (the prediction of the static model), but also lower than the upper boundary (the prediction of the simple iterative model). The average difference between the observed and predicted ASD by the simple iterative model was significant, (mean ± SD: -0.112±0.053, p<.0001; t-test, n=14), showing that the simple iterative model cannot adequately predict the serial dependence either, because it predicted a stronger serial dependence than the observed ASD.

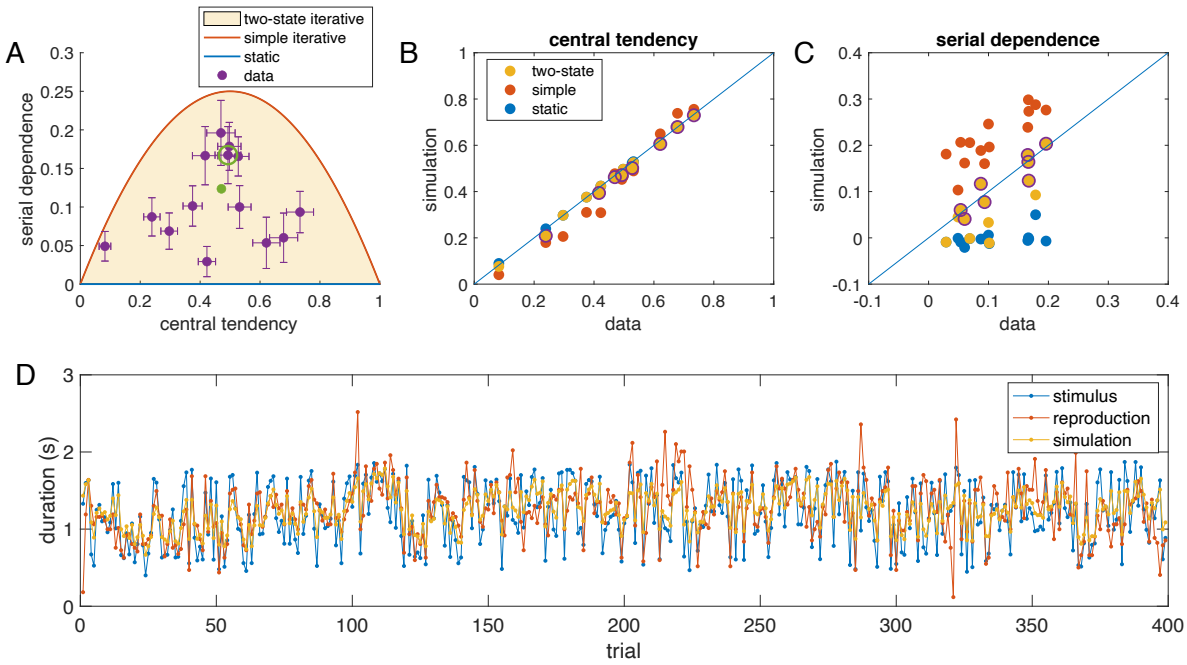

Figure 4: **A**. Absolute serial dependence in duration reproduction plotted over central tendency for all 14 subjects (purple dots, error bars show standard error, each point corresponds to 400 trials). The data point

corresponding to the raw data shown in Fig. 3 is marked by a green circle. Dots approximately lying on the red curve would be expected if the data can be described by the simple iterative model. Note that all data points are above the prediction for a static model (sequential effect of zero, blue line), but below the prediction for the simple iterative model (red curve). Error bars represent the standard error of the slopes of the linear fits corresponding to central tendency and serial dependence. The yellowish shaded area corresponds to range of values covered by the two-state model; the green point is the fit corresponding to the green circle. **B, C**: Comparison of experimental data (x-axis) and model simulation (y-axis) for central tendency (**B**) and absolute serial dependence (**C**) for all three models and all participants. Each dot represents one participant. The central tendency is captured correctly by all three models (values close to the diagonal), whereas the model values for serial dependence are too small for the static model (blue dots, close to zero), and too large for the simple iterative model (red dots). Values for the two-state model lie closer to the diagonal in most cases. Cases for which the two-state model was the best model according to cross-validation are highlighted by circles. The two-state model is appropriate independently of the size of central tendency or serial dependence for 8 of 14 participants, while in the remaining 6 cases the static model is preferable. **D**. Time course of stimulus (blue) and reproduction (red) from one participant (same data as in Fig. 3) together with best fit of the two-state model (yellow). The model fitting minimized the least-squares distance between reproduction (red) and simulation (yellow). The input to the model was the trial-by-trial stimulus time course (blue). The few outliers (e.g., around trial 105 or 320) were not removed for the model fitting.

We therefore fitted[†] the extended (two-state) iterative model to the individual data (3 free parameters per subject) to evaluate whether the model would capture not only the central tendency but also the serial dependence better than the simpler alternatives. An example time course of stimulus, data, and best-fit simulation is shown in Fig 4D (see also SI Appendix B: raw data in SI Fig. S3 and fits to mean responses in Figs. S4 and S5). Note that the fit minimizes the least-squares distance between the individual responses and the model simulation, which receives as input the trial-to-trial time course of the stimuli in exactly the same order as presented to the individual participant.

The static model and the simple iterative model are special cases of the two-state model: both are nested within the two-state model. Therefore, one can determine whether the parameters that are set to zero for the simpler models differ significantly from zero in the full model. On average, both parameters (Parameter 1: the relative variability of the stimulus distribution, and Parameter 2: the relative variability of the additive change of the mean, see Fig. 1A) of the full model were significantly different from zero (Parameter 1: 1.03±0.28, mean ± SEM, t-test $p<.01$; Parameter 2: 0.14±0.05, mean ± SEM, t-test $p=0.025$; both $n=14$). In individual participants, the relative variability of the stimulus distribution differed significantly from zero (assessed via confidence intervals of the parameters) for all subjects (range 0.20 to 4.12), while the variability of the additive change differed from zero only for 6 of 14 subjects (range 0 to 0.66). To determine which model was more appropriate for fitting the data, we used an out-of-sample cross validation procedure specifically suited for model selection in time series (Arlot & Celisse 2010). According to this cross-validation procedure (see Materials and Methods), the two-state model is the preferable model for 8 of 14 participants, while for the remaining 6 participants the static model is sufficient (Fig. 4). A comparison of the values for central

[†] Model simulation performed after transforming the stimulus data to the log-domain, as done previously (Petzschner & Glasauer 2011; Roach et al. 2017), to account for the dependence of the variance on stimulus size (Weber-Fechner law). Fitting was done by minimizing the least-squares distance between trial-wise response and simulation in the linear domain.

tendency and absolute serial dependence derived from the data and from respective model simulations is presented in Fig 4B and 4C. In case of perfect fit, all points should lie on the diagonal. While all three models capture well the central tendency (Fig. 4B), the ASD is too small for the static model but too large for the simple iterative model (Fig. 4C), while the two-state model matches the data reasonably well.

While the results so far confirm that the two-state model provides a quantitative explanation for central tendency and serial dependence at lag one (i.e., dependence on the previous stimulus), due to its iterative nature the two-state model also predicts dependence of the current error on stimuli further in the past. That this is indeed the case also experimentally can be shown by cross correlation analysis: for duration reproduction, the cross correlation between stimulus and reproduction is, on average, significantly different from zero up to lag 3 (t-test; lag 2: p=0.0007; lag 3: p=0.039; n=14; supplementary Fig. S4).

The averaged experimental results together with the averaged model results for dependence of error on current and previous stimuli is shown in Fig. 5 (see also SI Fig. S2). Note that the model was fitted to each individual trial-by-trial reproduction time course separately by minimizing the trial-wise least-squares distance between experimental reproduction and model simulation. Thus, the good match shown in Fig. 5B and 5C, quantified by a high coefficient of determination R^2, is caused by the model mimicking the experimental serial dependence without explicitly including it in the fitting procedure. It is not a trivial consequence of the model fit, as shown by the fact that both static and simple iterative models can fit the central tendency equally well (i.e., the dependence shown in Fig. 5A), but fail to correctly exhibit the serial dependence shown in Fig. 5B and 5C (see SI Appendix B1).

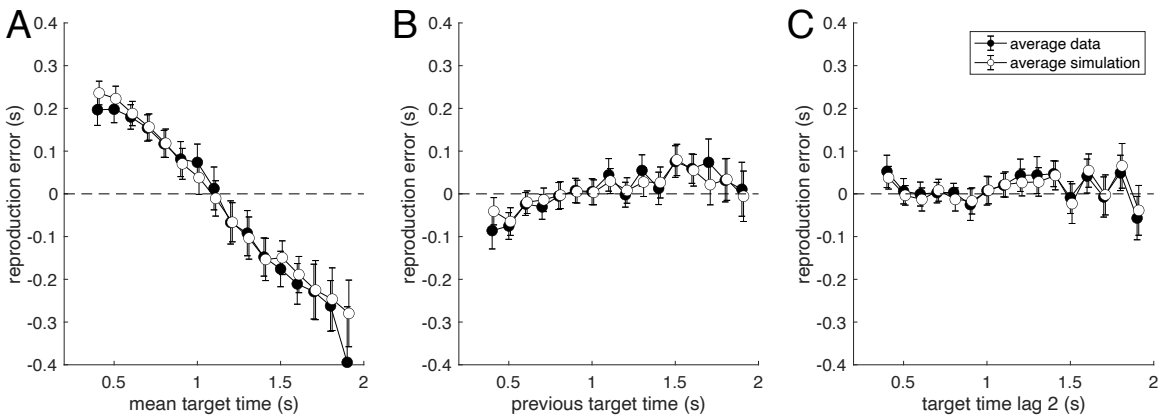

Figure 5: Average data for the duration reproduction experiment (black dots) together with average simulation result for the best fit iterative two-state model (open dots). Stimuli were presented in randomized order. **A**: Reproduction error as a function of the current target duration (R^2=0.964). **B**: Reproduction error as a function of the previous target duration (i.e., ASD, R^2=0.803). **C**: Reproduction error as a function of the target duration two trial in the past (R^2=0.834). Data points are averages of 7 to 14 subjects (only 7 subjects received long durations above 1.7 s); error bars denote SEM. For the model simulation, the model (three free parameters) was fitted individually to the trial-by-trial time course of the responses of each subject (same fits as for Figure 4). From theses individual simulated response time courses, the average simulation was computed in the same way as for the real response time courses. Note that not only average mean data and simulated responses match, but that also the size of the error bars is captured quite well by the model.

While the data analysed to far came from randomized stimuli, the experiment also included another condition, in which the same stimuli were presented in a random walk order to the same participants (see also Materials and Methods). In this condition, subsequent stimuli are similar to each other, just as supposed by the generative model of the simple iterative Bayes (see Fig. 2C, red time course, for an example of such a random walk). As explained in our previous paper (Glasauer & Shi 2021), this condition tests the prediction of the simple iterative and the Petzschner & Glasauer (2011) explanatory models, which both predict that the central tendency vanishes in the random walk condition.

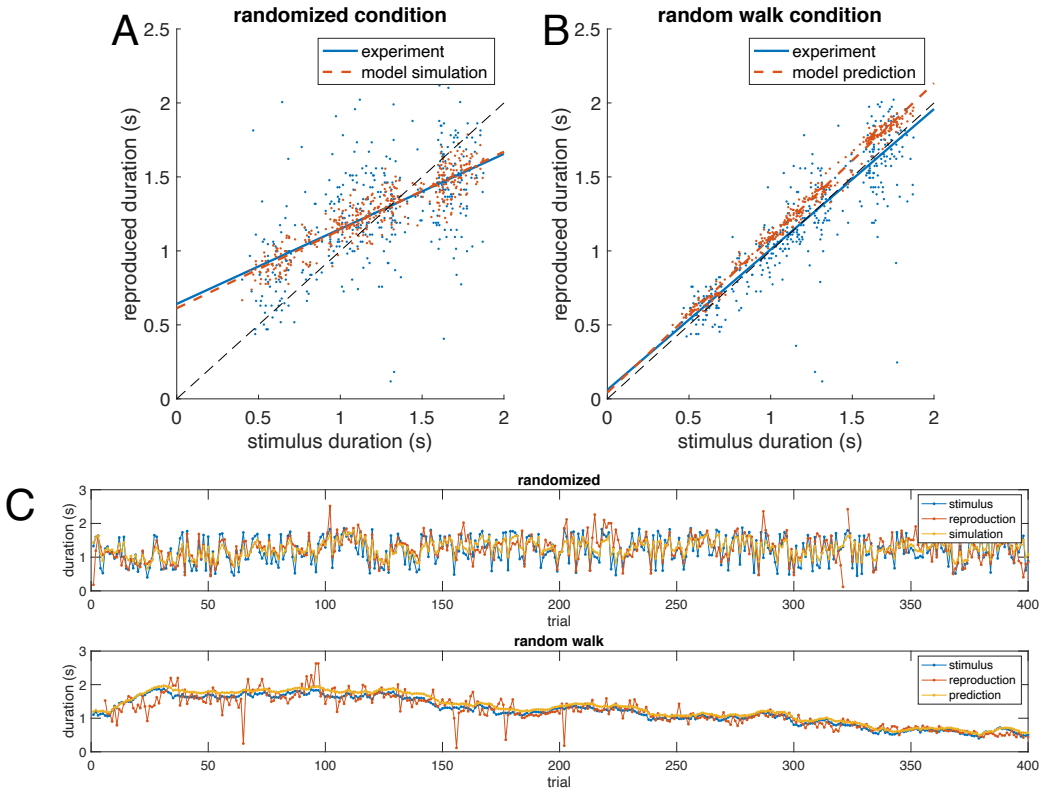

Figure 6: Example data from one participant and model results. **A**: Data (blue) and model simulation (best fit to two-state model, red) for the 'randomized' condition of the duration reproduction experiment. Lines are linear least squares fits to the data showing that the model replicates the central tendency. **B**: Data (blue) and model prediction (red, parameters same as in A) for the 'random walk' condition, in which successive stimuli were similar. It can clearly be seen that in this subject the central tendency dramatically changed, which was predicted by the model. **C**: Data from A and B plotted over trial. The same stimuli were used in both conditions.

However, while it was found that the central tendency indeed decreased substantially and was significantly smaller during random walk (t-test, n=14, $p<0.0001$; see Fig. 7A), it did not completely vanish and was still larger than predicted by these previous models (Glasauer & Shi 2021). This was, however, not true for all subjects: for some subjects, the central tendency was no longer different from zero (see example data in Fig. 6), while for others it clearly was still visible. Serial dependence also changed and became on average negative with a significant difference between conditions (Fig. 7B; t-test, n=14, $p=0.00016$).

Since we suspected that the remaining central tendency in the random walk condition and the change in serial dependence could be explained by the new two-state model introduced above, we used the individually fitted model parameters obtained from the randomized condition to predict the individual time courses of the random walk condition.

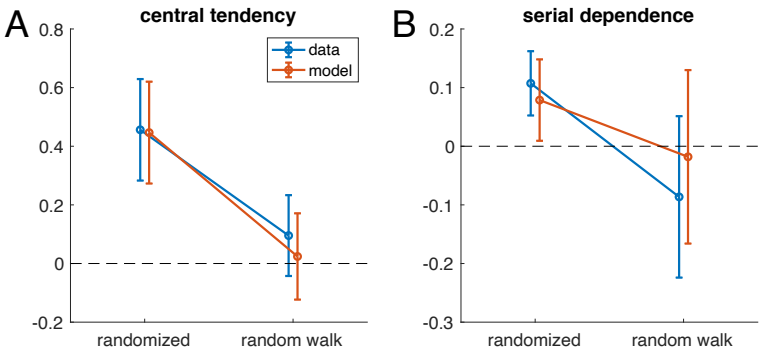

Figure 7: Summary of average results for both conditions of the duration reproduction experiment. Blue: data; orange: model. Error bars indicate standard deviation. **A:** Central tendency is significantly smaller during random walk (t-test, n=14, p<0.0001). **B**: Serial dependence becomes on average negative with a significant difference between conditions (t-test, n=14, p=0.00016). In A and B, for the random walk condition model predictions are shown using the parameters from fitting the randomized condition.

Figure 8 shows the averaged experimental results together with the averaged model prediction. Both central tendency (Fig. 8A) and serial dependence (Fig. 8B and 8C) are well-predicted by the model, showing that the central tendency remaining in the random walk condition is explained by the generative assumption of the two-state model. Note that the similarity of the error dependence on current and previous stimuli in the random walk condition shown in Fig. 8 is expected, since stimuli in this condition are highly autocorrelated, i.e., the current stimulus is indeed similar to the stimuli preceding it (and thus the reproduction error is similar when plotted over current or previous stimuli).

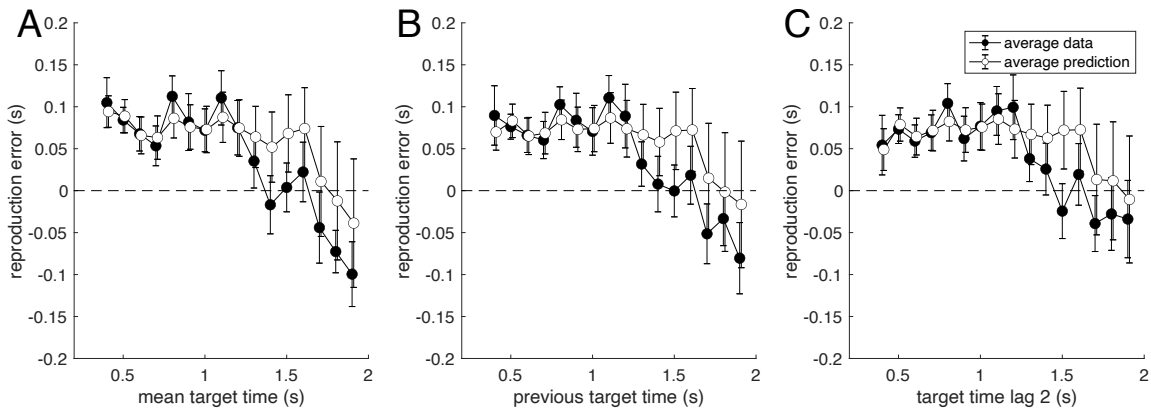

Figure 7: Average data for the 'random walk' condition of the duration reproduction experiment (black dots), in which successive stimuli were similar, presented together with averaged model predictions of the iterative two-state models used in Fig. 5 (open dots). Stimuli are the same as in Fig. 5 except for the order of presentation. Model parameters were the same as used in Fig. 5. **A**: Reproduction error as a function of the current target duration (R^2=0.629). **B**: Reproduction error as a function of the previous target duration (i.e., ASD, R^2=0.544). **C**: Reproduction error as a function of the target duration two trial in the past (R^2=0.428). Data points are averages of 7 to 14 subjects (only 7 subjects received long durations above 1.7 s); error bars denote SEM. Average model predictions are calculated from trial-by-trial predicted individual time courses using the parameters corresponding to Fig. 5.

We also analysed the publicly available data set (Petzschner & Glasauer 2020) published previously (Petzschner & Glasauer 2011), using the same method. The data come from two separate experiments on visual path integration, one on linear distance reproduction and one on

reproduction of angular distance (see Materials and Methods). While Petzschner and Glasauer (2011) showed that their iterative model could well capture the central tendency, they did not analyse serial dependence. Fig. 8 shows the equivalent analysis as above for the two path-integration experiments (Petzschner & Glasauer 2011). For the linear distance reproduction, the average serial dependence is 0.100±0.045 (mean ± SD). For the angular distance reproduction, it is 0.119±0.057 (mean ± SD). As for duration reproduction, the data for the two distance reproduction experiments confirms that neither the static nor the simple iterative model can capture the serial dependence sufficiently well (all *p*s<0.0001).

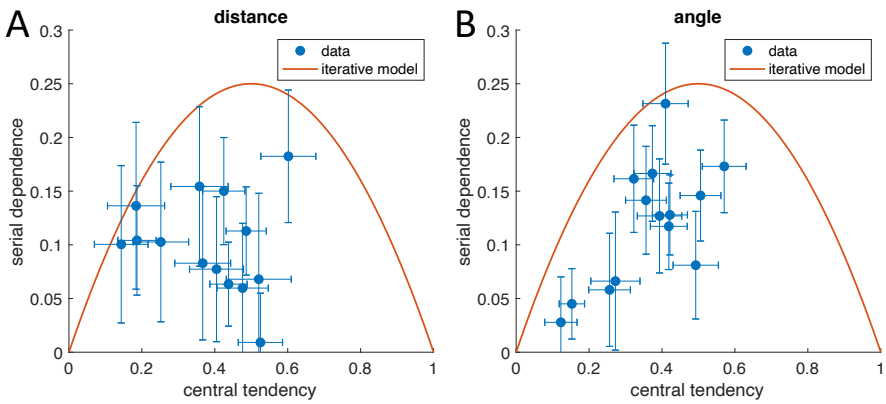

Figure 8: The absolute serial dependence for distance (**A**) and angular (**B**) reproduction plotted over central tendency for all subjects. Distance and angular reproduction data come from Petzschner & Glasauer 2011, data published as Petzschner & Glasauer 2020, each point corresponds to 540 trials. The same analysis was applied as in Fig. 4A.

The results of fitting the two-state model to the distance reproduction experiments are shown in SI Appendix B2. In both cases, the sequential dependence is predicted extremely well by the two-state model.

Discussion

In this paper, we analysed the relation between the central tendency and serial dependence for magnitude reproduction with two aims: to distinguish between static and iterative models proposed in the literature to explain the central tendency bias, and to reveal the origin of individual differences seen experimentally. We analysed three datasets, one from duration reproduction and two from path integration, to evaluate which model can better explain magnitude reproduction regarding both the serial dependence, quantified as ASD (current error depending on previous stimulus magnitude), and the central tendency effects. Effects of immediate prior experience on current decisions have been reported for various cases in the psychological literature, and, as we show here, they are also clearly visible in experiments on magnitude reproduction. The average ASDs found in duration and distance reproduction differed significantly from zero, which clearly demonstrates that the response error of the current trial depends on the stimulus from the last trial. This contradicts static models which imply no influence of the previous trial. Consequently, several previously published static models can be ruled out (Jazayeri & Shadlen 2010, Roach et al. 2017, Lakshminarasimhan et al. 2018). Even though the static models fit the central tendency well in experiments with random stimulus presentation, the explanation for this bias offered by the static models is only

partially correct. The fundamental assumption of the static models, a fusion of the sensed stimulus with prior information about the stimulus range, is not completely wrong, except that the prior is not static, as shown by the significant serial dependence. Rather, the prior is updated trial-by-trial so that information from the immediate previous trial is used for the current estimate. Due to the iterative nature of Bayesian estimation with the posterior as basis for the new prior, not just the previous stimulus (as proposed by Cicchini et al. 2018), but also stimuli further in the past can still exert an influence on the present response. This difference between static and iterative models has important consequences for understanding the processes that lead to the perceptual results: while the results of static and iterative models look similar with respect to the central tendency, the internally represented priors, the underlying generative model assumptions, and the predictions for the sequential dependence are completely different.

Like the static prior model, the simple iterative model used previously (e.g., Dyjas et al. 2012; Glasauer 2019), predicts the central tendency effect very well but falls short in accounting for the serial dependence. The simple iterative model assumes that stimuli remain the same from trial to trial except for a random change. This generative assumption corresponds to stimuli being generated by a random walk or discrete Wiener process. According to this assumption, the overall variance of the stimuli builds up over the trials. By contrast, the static model assumes that the stimulus distribution has a fixed variance, and a fixed mean. The generative assumption for the iterative model also implies a stimulus sequence that differs considerably from that of the static model: it resembles Brownian motion or a diffusion process in one dimension rather than a random sequence (see Fig. 2C for examples). Both the static and the simple iterative models provide predictions concerning the serial dependence: the static model predicts zero serial dependence (if quantified as ASD, blue line in Fig. 4), the iterative model predicts that, in case of random stimuli, serial dependence depends on central tendency in a predictable way (see red curve in Figs. 4 and 7).

The empirical data, however, showed that neither of these two models captures the experimental relation between central tendency and serial dependence. The two-state model, combining the static and the simple iterative models, assumes that stimuli at each trial come from a distribution with fixed variance, but that the mean of that distribution changes from trial to trial. By merging the generative assumptions of static and simple iterative models, both the central tendency effect and the absolute serial dependence can be well explained. According to the two-state model, the considerable variations between participants are not only caused by different impact of noise on sensory measurement, but also because of different beliefs concerning the sequential structure of the stimuli. As an example, in Fig. 4A, of two participants with approximately the same central tendency of 0.42, one had a serial dependence of 0.03, the other of 0.17. This difference reflects the observers' own supposition about the sequential structure: the participant with a low serial dependence assumed the world is volatile and trusted only the current stimulus together with a hypothesis about the limited range of stimuli for perceptual estimates. By contrast, the participant with a large serial dependence agreed about the randomness of the world but further assumed that things change over time with some continuity. The present investigation also suggests that an observer's belief about the world's sequential structure is carried over from one experimental condition to another instead of being adapted to an

individual condition: the model parameters derived from the randomized condition of duration reproduction provided an excellent prediction of the experimental results of the random walk condition, even though both conditions varied exactly (and only) by their sequential structure. However, whether these beliefs reflect intrinsic personality traits warrants further investigation.

Another question is whether the two-state model can encompass the full spectrum of empirical values for central tendency and serial dependence. The two-state model predicts that serial dependence quantified as ASD should, with randomized stimuli, not exceed the value predicted by the simple iterative model, that is, the quadratic relationship with central tendency (red curve in Figs. 4 and 7), which has a maximum at 0.25. Indeed, this is the case for the three experiments we validated here. Note, however, that this is not a trivial result: for example, a model proposed previously to explain serial dependence in visual orientation reproduction (Cicchini et al. 2018) predicts serial dependence that is approximately equal to central tendency and can assume values as large as 0.5 (see SI Appendix C). Evidently, this model cannot explain the present data, but it shows that there are alternatives to the two-state model, which would allow serial dependence larger than 0.25. However, in our experiments, serial dependence did not, for any of the tested participants, exceed the theoretical maximum postulated by the two-state model, which again suggests that our model provides a good explanation for the participants' behaviour.

We also showed that a commonly used relative measure of sequential dependence, which uses current error depending on the difference between current and previous stimulus (the RSD), is not suitable for quantification, if stimuli come from an open scale, such as duration or distance. In this case, the RSD shows a bogus dependence on previous trials because of mathematical coupling (see SI Appendix A1). While most studies using the RSD employed it correctly to circular stimuli, at least one recent study investigated both central tendency and serial dependence for facial age estimation (Clifford et al. 2018), but unfortunately their conclusions were based on the RSD quantification that does not hold for age, which is a stimulus on an open scale.

One might wonder about the purpose of integrating immediate prior information into a current decision, given that it may cause an estimation bias. One common explanation is that the regularity of our environment is relatively stable, so that integrating prior knowledge will boost the reliability of the estimation and facilitate performance (Petzschner et al. 2015, Shi et al. 2013). For a visual orientation reproduction task (Cicchini et al. 2018), the authors argued that serial dependence provides a behavioural advantage manifesting with low reaction times and high accuracy. When the stimuli are similar between trials, it is useful to use the last perceived stimulus as prior. This assumption about the sequential structure is included in the generative assumption of the two-state model: the stimulus of the current trial is assumed to be similar to that of the last trial, since it comes from a distribution with a similar mean. However, the mean of the sampled stimuli also fluctuates over time, which makes the two-state model more flexible than a static model. That is, observers do not assume that the randomness of the external environment is strictly stable, but rather expect variations and changes.

Next, the question arises whether the proposed two-state model is appropriate or optimal for the usual experimental situations. The answer is obvious: for the experiments investigated here, in which stimuli are randomly drawn from a fixed distribution, it is not. In the experimental paradigm, which has been used over and over since Vierordt's work in 1868, stimuli are randomly generated from a fixed, pre-defined distribution, which makes including a sequential dependence unnecessary. Using the last trial to estimate the current would then deteriorate rather than improve the quality of the estimate. However, as evidenced by the significant serial dependence, most our participants assume that there is at least some temporal continuity in the stimuli. According to the model, for these participants, the overall central tendency bias should be smaller, if the stimuli are indeed similar from trial to trial. This was validated by showing in our previous study (Glasauer & Shi 2021), that the central tendency in sequences with complete random stimulus order was larger than in sequences with random-walk fluctuation. Here we showed that this decrease in central tendency and, more importantly, the remaining central tendency, is well-predicted by the two-state model.

Finally, our results show that the individual differences that are expressed in different values of central tendency and serial dependence are not only due to differences in sensory noise, but reflect major differences in the underlying generative model, that is, in the assumptions about how stimuli are generated in the world. While some participants behave as if stimuli are generated almost independently from each other, just like when sampling from a random distribution, others show strong serial dependence and thus assume that subsequent stimuli are similar in size and depend on each other. While on average, the perceptual system of participants seems to be optimized for random stimuli the distribution of which slowly changes over the time, the individual differences in belief about stimulus generation are not negligible.

In summary, our two-state iterative model assumes that the magnitude percept is an integration of sensory input and an updating internal prior, the mean of which changes from trial to trial with some fluctuation. The model explains the variable link between the sequential dependence and the central tendency in the generative assumptions about the sequence structure, which differs among participants. It thus allows not only modelling the average responses of participants but also elucidates the reason for their variability: the assumptions behind the perceptual estimation process vary from person to person. The same world looks different for each of us, even when considering such a basic ability as perceiving magnitudes.

Materials and Methods

Duration reproduction

14 naïve volunteers (7 female, average age 27.4) participated in the experiment, which was approved by the ethics committee of the Department of Psychology at LMU Munich. A yellow disk (diameter 4.7°, 21.7 cd/m2) was presented as visual stimulus on a 21-inch monitor (100 Hz refresh rate) at 62 cm viewing distance using the Psychtoolbox (http://psychtoolbox.org). Each trial started after 500 ms presentation of a fixation cross followed by the stimulus which appeared for a pre-defined duration. After a short break of 500 ms participants were prompted

to reproduce the duration of the stimulus by pressing and holding a key. The visual stimulus was shown again during key press. At the end of the trial, a coarse visual feedback was given for 500 ms (5 categories from <-30% to >30% error). Each participant performed two blocked sessions in balanced order. In the random walk condition, participants received 400 stimuli generated by cumulative summation (integration) of randomly distributed values from a normal distribution with zero mean and a SD that was chosen to yield stimuli between 400 ms and 1900 ms. In the randomized condition, the same 400 stimuli were used in scrambled order. Each participant received a different sequence (see Fig. 5D for an example). The data have been used previously in Glasauer & Shi (2021).

Distance reproduction

The procedure has been published previously (Petzschner & Glasauer 2011) and the data are publicly accessible (Petzschner & Glasauer 2020). Briefly, 14 volunteers (7 female, aged 22–34 years) participated. Stimuli were presented in darkness on a computer monitor as real-time virtual reality using Vizard 3.0 (Worldviz) depicting an artificial stone desert consisting of a textured ground plane, 200 scattered stones placed randomly, and a textured sky. Participants used a joystick to navigate. Estimation of travelled distances and of turning angles was tested separately under three different conditions (different ranges of distances or angles, see Fig. S5 and S6, 200 trials per condition) in a production-reproduction task. For distance estimation, participants were instructed to move forward on a linear path until movement was stopped when reaching the randomly selected production distance (same sequence for all subjects) and then had to reproduce the perceived distance in the same direction using the joystick and indicate their final position via button press. Velocity was kept constant during movement but randomized up to up to 60% to exclude time estimation strategies. No feedback was given. For angular turning estimation, the procedure was the same except that subjects had to turn.

Data analysis: central tendency and serial dependence

To quantify central tendency, a linear least-squares regression was fitted to stimulus reproduction plotted over stimulus duration for each participant individually using Matlab (The Mathworks, Natick MA, USA). Central tendency was defined as 1-slope of the regression line. Absolute serial dependence (ASD) was assessed by fitting a linear least-squares regression to the error in trial k plotted over the stimulus in trial k-1.

Modelling: The two-state model

The generative equations for the two-state model are given as (see Fig. 3 for schematic illustration):

$$x_i = m_{i-1} + \varepsilon_x$$
$$m_i = m_{i-1} + \varepsilon_m$$
$$z_i = x_i + \eta$$

with x_i being the stimulus at trial i that is drawn from a distribution with mean m_{i-1} and variance v (here expressed by the random number ε_x, which is normally distributed as $N(0, v)$).

The mean of this stimulus distribution m_i at the trial i is the same as in the trial before except for the random fluctuation ε_m (ε_m is normally distributed as $N(0, q)$). The actual sensory measurement (or sensation) z_i is the stimulus corrupted by the sensory noise η, which is normally distributed as $N(0, r)$.

We can rewrite these equations in matrix notation with $X_i = \begin{bmatrix} x_i \\ m_i \end{bmatrix}$, so that

$$X_i = FX_{i-1} + \varepsilon$$
$$z_i = H \cdot X_i + \eta$$
$$F = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$
$$H = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

The optimal estimator for this model can be written as time-discrete Kalman filter:

$$\hat{X}_{k|k-1} = F\hat{X}_{k-1}$$
$$P_{k|k-1} = FP_{k-1}F^T + Q$$
$$K_k = P_{k|k-1}H^T\left(HP_{k|k-1}H^T + r\right)^{-1}$$
$$\hat{X}_k = \hat{X}_{k|k-1} + K_k\left(z_k - H\hat{X}_{k|k-1}\right)$$
$$P_k = (I - K_kH)P_{k|k-1}$$

The steady state with constant matrix K thus becomes

$$\hat{X}_k = F\hat{X}_{k-1} + K_k\left(z_k - HF\hat{X}_{k-1}\right)$$

Written as states x and m, this can be expressed as

$$\hat{x}_i = \hat{m}_{i-1} + k_1(z_i - \hat{m}_{i-1})$$
$$\hat{m}_i = \hat{m}_{i-1} + k_2(z_i - \hat{m}_{i-1})$$

Free parameters of the model so far are the variance ratios v/r and q/r.

It should be noted that the model operates in log space (Petzschner & Glasauer 2011; Roach et al. 2017) to account for the Weber-Fechner law. Raw sensory input d_i is thus transformed logarithmically to yield z_i with $z_i = \ln\left(1 + \frac{d_i}{d_0}\right)$. The stimulus estimate $\hat{x}_i$ is finally back-transformed to yield $\hat{d}_i$ with $\hat{d}_i = d_0 \cdot e^{\hat{x}_i + \Delta x}$. The shift term Δx accounts for possible choices of the cost function and is the third free parameter of the full model.

To summarize, the two-state model has three free parameters:

1) the ratio of the variances v and r indicating the variability of the stimulus distribution relative to the sensory noise,
2) the ratio of variances q and r indicating the variability of the additive random shift relative to the sensory noise, and
3) a shift parameter that accounts for global over- or underestimation (see also Petzschner & Glasauer 2011).

For all three models (static, simple iterative, two-state) the same model equations and the same Kalman filter can be applied. The three models differ by the free parameters:

1) static model: $\varepsilon_m = 0$, therefore variability $q = 0$. Free parameters: v/r and Δx.
2) iterative model: $\varepsilon_x = 0$, therefore variability $v = 0$. Free parameters: q/r and Δx.
3) two-state model: full model. Free parameters: q/r, v/r, and Δx.

Model fitting and model selection

For model simulation, the individual stimulus sequences were used to fit the model separately for each participant. Thus, the model received the sequence of stimuli in exactly the same order as the participant and computed a sequence of responses. Model fitting was performed in linear stimulus space, that is, for model fitting, the least-squares distance between stimulus sequence and responses was minimized. The Matlab function *lsqnonlin* was used to estimate the parameters, and *nlparci* was applied to estimated confidence intervals.
The coefficient of determination R^2 for average data was calculated as $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, with SS_{res} being the residual sum-of-squares and SS_{tot} the total sum-of-squares. If a model perfectly captures the data, $R^2 = 1$. Models with negative R^2 are worse than the baseline model, which predicts the average of the data and which has $R^2 = 0$.

To compare models, we used a leave-one-out (LOO) cross-validation procedure (Arlot & Celisse 2010) adapted to time series. In LOO, each of the n data points is successively "left out" from the sample and used for validation by fitting the model to the remaining data points and recording the error of the left-out data point. The criterion is the average validation error of the n model fits: the best model is the one with the minimal validation error. To account for the trial-to-trial dependence of data or model, in the modified LOO instead of leaving out just one data point, k values around this data point are left out, while the validation error is only computed for the data point in the center of this leave-out window. Here we selected k=11, which was assumed to be large enough to account for trial-to-trial dependencies. The same result (8x two-state model selected) was already achieved with k=3, while for k=1 the two-state model was best in 9 cases.

Acknowledgements

This work was supported by German Research Foundation (DFG) grants GL 342/3-2 and SH 166/3-2. We thank Mauro Manassi, David Whitney, and Jason Fischer for pointing out that for stimuli on a circular scale the RSD is valid without causing artifacts, and that for the RSD permutation tests should be run as statistical sanity check.

References

1. Archie, J. P. Mathematical coupling: a common source of error. *Ann. Surg.* **193**, 296–303 (1981)
2. Arlot, S. & Celisse, A. Cross-validation procedures for model selection. *Stat. Surv.* **4***,* 40-79 (2010)
3. Bausenhart, K. M., Dyjas, O. & Ulrich, R. Temporal reproductions are influenced by an internal reference: Explaining the Vierordt effect. *Acta Psychol* **147***,* 60–67 (2014)
4. Bliss, D. P., Sun, J. J., D'Esposito, M. Serial dependence is absent at the time of perception but increases in visual working memory. *Sci Rep* **7**, 14739 (2017)

5. Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M. & Burr, D. C. Optimal encoding of interval timing in expert percussionists. *J. Neurosci.* **32**, 1056–1060 (2012)
6. Cicchini, G. M., Mikellidou, K., & Burr, D. C. The functional role of serial dependence. *Proc. Biol. Sci.* **285**, 20181722 (2018)
7. Clifford, C. W. G., Watson, T. L. & White, D. Two sources of bias explain errors in facial age estimation. *R. Soc. Open Sci.* **5***,* 180841 (2018)
8. Cross, D. V. Sequential dependencies and regression in psychophysical judgments. *Percept. Psychophys.* **14**, 547-552 (1973).
9. Curran-Everett, D. Explorations in statistics: correlation. *Adv. Physiol. Educ.* **34**, 186 –191 (2010)
10. Dyjas, O., Bausenhart, K. M. & Ulrich, R. Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. *Atten. Percept. Psychophys.* ***74,*** 1819–1841 (2012)
11. Ernst, M. O. & Bülthoff, H. H. Merging the senses into a robust percept. *Trends Cogn. Sci.* **8,**162-169 (2004)
12. Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nat. Neurosci.* **17**, 738-743 (2014)
13. Glasauer, S. & Shi, Z. "Central tendency as consequence of experimental protocol" in *Proceedings of the 2019 Conference on Cognitive Computational Neuroscience*, (2019), pp. 268-271. doi: 10.32470/CCN.2019.1148-0
14. Glasauer, S. & Shi, Z. The origin of Vierordt's law: the experimental protocol matters. *PsyCh J* **10**, (2021) https://doi.org/10.1002/pchj.464
15. Glasauer, S. Sequential Bayesian updating as a model for human perception. *Prog. Brain. Res.* **249**, 3-18 (2019)
16. Holland, M. K. & Lockhead, G. R. Sequential effects in absolute judgments of loudness. *Percept. Psychophys.* **3***,* 409–414 (1968).
17. Hollingworth, H. L. The central tendency of judgment. *J. Phil. Psychol. Sci. Meth.* **7***,* 461-469 (1910).
18. Jazayeri, M. & Shadlen, M. N. Temporal context calibrates interval timing. *Nat. Neurosci.* **13***,* 1020-1026 (2010)
19. Kiyonaga, A., Scimeca, J. M., Bliss, D. P. & Whitney D. Serial dependence across perception, attention, and memory. *Trends Cogn. Sci.* **21**, 493-497 (2017)
20. Lakshminarasimhan, K. J., et al. A dynamic Bayesian observer model reveals origins of bias in visual path integration. *Neuron* **99**, 194-206 (2018)
21. Liberman, A., Fischer, J. & Whitney, D. Serial dependence in the perception of faces. *Curr. Biol.* **24**, 2569-2574 (2014)
22. Petzschner, F. & Glasauer, S. Magnitude estimation data. G-Node https://doi.org/10.12751/g-node.21796b (2020)
23. Petzschner, F. H. & Glasauer, S. Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *J. Neurosci.* **31***,* 17220–17229 (2011)
24. Petzschner, F. H., Glasauer, S. & Stephan, K. E. A Bayesian perspective on magnitude estimation. *Trends Cogn. Sci.* **19**, 285-293 (2015)

25. Powell, G., Meredith, Z., McMillin, R. & Freeman T. C. Bayesian models of individual differences. *Psychol. Sci.* **27**, 1562-1572 (2016)
26. Roach, N. W., McGraw, P. V., Whitaker, D. J. & Heron, J. Generalization of prior information for rapid Bayesian time estimation. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 412-417 (2017)
27. Shi, Z. & Burr, D. Predictive coding of multisensory timing. *Curr Opin. Behav. Sci.* **8**, 200–206 (2016)
28. Shi, Z., Church, R. M. & Meck, W. H. Bayesian optimization of time perception. *Trends Cogn. Sci.* **17**, 556-564 (2013)
29. Thurley, K. Magnitude estimation with noisy integrators linked by an adaptive reference. *Front. Integr. Neurosci*. **10**, 6 (2016)
30. Vierordt, K. *Der Zeitsinn nach Versuchen.* (H. Laupp'sche Buchhandlung, Tübingen, 1868).