Magnitude Estimation with Noisy Integrators Linked by an Adaptive Reference

Kay Thurley [1,2*]

[1] Department Biology II, Ludwig-Maximilians-Universität München, München, Germany, [2] Bernstein Center for Computational Neuroscience, Munich, Germany

Judgments of physical stimuli show characteristic biases; relatively small stimuli are overestimated whereas relatively large stimuli are underestimated (regression effect). Such biases likely result from a strategy that seeks to minimize errors given noisy estimates about stimuli that itself are drawn from a distribution, i.e., the statistics of the environment. While being conceptually well described, it is unclear how such a strategy could be implemented neurally. The present paper aims toward answering this question. A theoretical approach is introduced that describes magnitude estimation as two successive stages of noisy (neural) integration. Both stages are linked by a reference memory that is updated with every new stimulus. The model reproduces the behavioral characteristics of magnitude estimation and makes several experimentally testable predictions. Moreover, the model identifies the regression effect as a means of minimizing estimation errors and explains how this optimality strategy depends on the subject's discrimination abilities and on the stimulus statistics. The latter influence predicts another property of magnitude estimation, the so-called range effect. Beyond being successful in describing decision-making, the present work suggests that noisy integration may also be important in processing magnitudes.

Keywords: magnitude estimation, interval timing, drift-diffusion model, uncertainty, regression effect, range effect, optimality

1. INTRODUCTION

In daily life we continuously need to process the physical conditions of our environment; we make judgements about the magnitude of sensory stimuli, represent them neurally and base decisions upon them. Judgements about magnitudes are inherently unreliable due to noise from different sources such as the statistics of the physical world, the judgement process itself, the neural representation of the stimulus and finally the computations that drive behavior. A large body of experimental work highlights that magnitude estimation is subject to characteristic psychophysical effects. These effects are strikingly similar across different sensory modalities, suggesting common processing mechanisms that are shared by different sensory systems (for a recent review see Petzschner et al., 2015). Amongst the behavioral characteristics the most astonishing yet unresolved is the *regression effect* also known as regression to the mean, central tendency, or Vierordt's law (von Vierordt, 1868; Hollingworth, 1910; Shi et al., 2013). It states that over a range of stimuli, small stimuli are overestimated whereas large stimuli are underestimated (**Figure 1A**). Regression becomes more pronounced for ranges that comprise larger stimulus

FIGURE 1 | Psychophysical characteristics of magnitude estimation. The typical properties of magnitude estimation are illustrated as they are reproduced by the model presented in this paper. The description is based on subsecond interval timing (cf., Jazayeri and Shadlen, 2010). **(A)** Individual reproduced values for each trial and stimulus (small dots, 100 per stimulus value), and their averages (large circles connected by lines) are shown for a simulation with three stimulus ranges. The regression effect is the deviation of the averages from the line of equality (diagonal gray dashed line) toward the mean of the respective stimulus range. It becomes stronger with larger means of the stimulus range, i.e., range effect. The analytical approximation of the model is in line with the simulated data (black solid lines). The memory parameter a was chosen to minimize MSE_r for each range (derived in Section "3.1"). Stimulus ranges and memory weights a are given in the top-left corner of the plot. Other parameters are $A_m = A_r = 0.25$, $\sigma_m = 1$, and $\sigma_r = 0.5$. Inset: Average deviations (BIAS) from the line of equality for each stimulus and test range. Solid lines are again analytical predictions. **(B)** Standard deviation and coefficient of variation (standard deviation divided by the mean) corresponding to **(A)**. Black solid lines are again analytical predictions. **(C)** Sequential effects. Plotting the response bias for a certain stimulus as a function of the stimulus in the previous trial, reveals effects of stimulus order in the simulations (thick lines). The simulation results can be analytically approximated (thin lines). Results for the range $494 - 847$ ms are displayed. For each stimulus value 10,000 trials were simulated.

values (*range effect*; Teghtsoonian and Teghtsoonian, 1978). As a consequence the same stimuli lead to different responses on average when embedded in different but overlapping stimulus distributions (**Figure 1A**) — the responses depend on the stimulus context (Jazayeri and Shadlen, 2010). Another omnipresent effect in magnitude estimation experiments is *scalar*

variability, i.e., errors monotonically increase with the size of the stimulus, attributed to the famous Weber-Fechner law **Figure 1B**; (Weber, 1851; Fechner, 1860). Finally, magnitude estimation is influenced by the sequence in which stimuli are presented (Cross, 1973; Hellström, 2003; Dyjas et al., 2012). According to such *sequential effects* the estimate of the stimulus in a particular trial is affected by the previous trial. This results in under- or overestimation of the current stimulus depending on the previous stimulus (**Figure 1C**).

The above behavioral characteristics likely result from an optimal strategy when noisy estimates are made about stimuli that itself depend on the statistics of the environment. Recently such optimality strategies were successfully explained in Bayesian frameworks (Jazayeri and Shadlen, 2010; Petzschner and Glasauer, 2011; Cicchini et al., 2012). Bayesian models incorporate a-priori knowledge about the stimuli into the estimation process, which seems to be crucial in explaining the aforementioned behavioral phenomena. However, the cited Bayesian approaches represent conceptual descriptions; inference about brain implementation is challenging.

The present paper introduces a theoretical approach that formulates magnitude estimation with noisy integrators (drift-diffusion processes). The model comprises two successive stages, measurement and reproduction. During measurement the current stimulus is estimated via noisy integration. The estimate is then combined with information from previous trials and used as threshold in the reproduction stage. The first passage of the threshold during reproduction determines the magnitude of the reproduced stimulus. Since the threshold depends on both the current and previous trials, it acts as an internal reference memory that is updated with every new stimulus. As we will see below, the model reproduces the behavioral characteristics of magnitude estimation (**Figure 1** anticipates these results) and interprets them as a consequence of an optimization strategy to minimize reproduction errors given noisy estimates and stimulus statistics.

2. MATERIALS AND METHODS

The analytical methods employed in this paper rely on standard mathematical and statistical techniques. Simulation and numerical analysis was performed with Python 2.7 using the packages: Numpy 1.9, Scipy 0.15, Statsmodels 0.6 (Seabold and Perktold, 2010), and Matplotlib 1.4 (Hunter, 2007). The model's stochastic differential equations, Equations (1, 4), were simulated via the approximation

$$x_{i+1} = x_i + A\,\Delta t + \sigma\,\sqrt{\Delta t}\cdot\mathcal{N}(0,1)\,.$$

A time step $\Delta t = 5$ ms was used, to appropriately sample the Gaussian process $\mathcal{N}(0,1)$, and capture noise sources on fast time scales like sensory noise and irregular spiking dynamics, at reasonable computing times.

2.1. Definition of the Model
Estimating the magnitude of a stimulus comprises two stages: First the stimulus is measured and afterwards the measurement

is reported, e.g., reproduced by matching the strength of the stimulus. In the present paper, both measurement and reproduction are modeled as drift-diffusion processes (e.g., Bogacz et al., 2006). During measurement drift-diffusion is left running as long as the stimulus is presented. Whereas, in the reproduction stage the drift-diffusion process is not stopped from outside but lasts until it hits a threshold from below. This threshold depends on the stimulus as estimated in the measurement stage and also includes the history of thresholds from previous trials, serving as an internal reference. **Figure 2** gives an overview of the model. For simplicity, the description below focuses on the estimation of temporal intervals (*interval timing*; Merchant et al., 2013). Numbers refer to interval timing in the subsecond range after Jazayeri and Shadlen (2010). However, application to the estimation of, e.g., sound intensity or spatial distances, is straightforward by reinterpreting the variables accordingly.

2.1.1. Measurement

The measurement stage comprises a drift process with rate A_m that is corrupted by noise (diffusion) realized as a Wiener process W with an amplitude σ_m. The dynamics are described by a stochastic differential equation

$$dm = A_m \, dt + \sigma_m \, dW, \quad m(0) = 0 \,. \tag{1}$$

The process is assumed to finish with the end of the stimulus and its final state yields the measurement. We can calculate the latter by integrating the above formula between stimulus start at $t = 0$ and end at $t = T$ (Broderick et al., 2009) and obtain

$$m_T := m(T) = A_m \, T + \sigma_m \sqrt{T} \cdot \mathcal{N}(0, 1) \,. \tag{2}$$

For convenience let us write m_T when we are considering a trial in which the interval T was presented, i.e., $m(T)$. The final

value m_T of the measurement process is Gauss-distributed $m_T \sim \mathcal{N}(A_m T, \sigma_m^2 T)$ with mean $\bar{m}_T = E(m \mid T) = A_m T$ and variance $\text{Var}(m_T) = \text{Var}(m \mid T) = \sigma_m^2 T$. This value is incorporated into the threshold of the reproduction phase as detailed below.

For later use, let us also derive the overall variance of the measurement $\text{Var}(m)$ here. To calculate $\text{Var}(m)$, we apply the law of total variance and get

$$\begin{aligned} \text{Var}(m) &= E(\text{Var}(m \mid T)) + \text{Var}(E(m \mid T)) \\ &= E\left(\sigma_m^2 \, T\right) + \text{Var}(A_m \, T) \\ &= \sigma_m^2 \, E(T) + A_m^2 \, \text{Var}(T) \,. \end{aligned} \tag{3}$$

2.1.2. Reproduction

Similarly to the measurement stage, reproduction is modeled as drift-diffusion with corresponding drift A_r and noise amplitude σ_r. However, here, the process is not stopped after a certain time but limited by an upper bound, i.e., a threshold θ (Broderick et al., 2009),

$$dr = A_r \, dt + \sigma_r \, dW, \quad r(0) = 0 \text{ and } r(t) < \theta \,. \tag{4}$$

The time of threshold crossing from below, i.e., the first-passage time of the drift-diffusion process, represents the response or the reproduced stimulus interval, respectively. Since we have a drift-diffusion process with a single threshold $\theta > 0$, the distribution of its first-passage times has an inverse Gaussian density $IG(\mu, \lambda)$: $f(x; \mu, \lambda) = \left[\frac{\lambda}{2\pi x^3}\right]^{1/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)$ and is characterized by $X \sim IG(\mu, \lambda) : E(X) = \mu, \text{Var}(X) = \mu^3/\lambda$ (Tuckwell, 1988). In
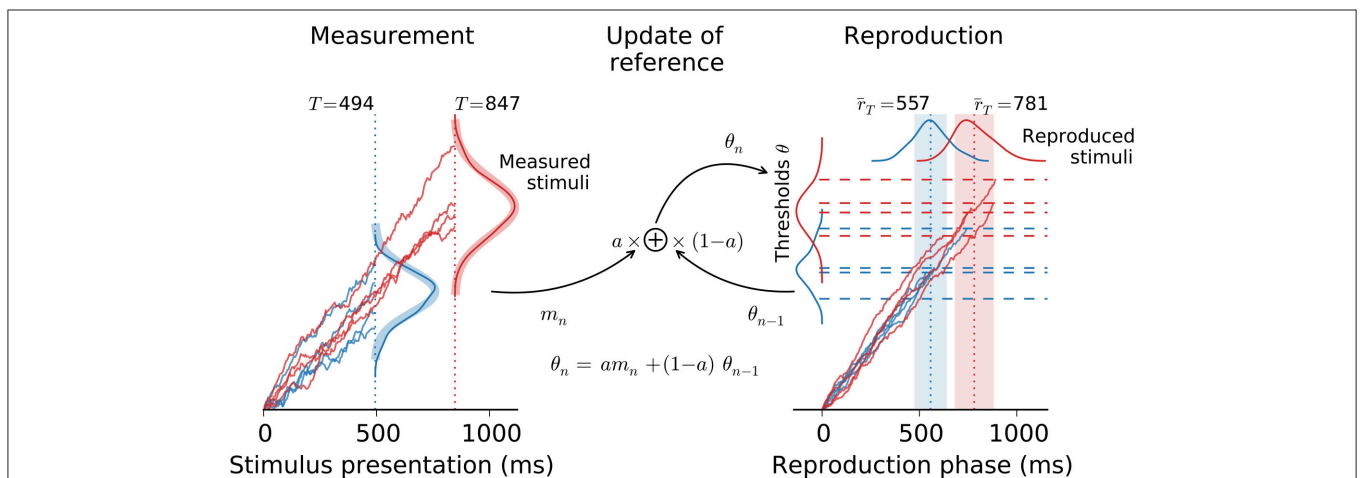

FIGURE 2 | Architecture of the model. The model comprises the measurement of the stimulus followed by its reproduction. Both stages are connected via the threshold θ for the reproduction stage (dashed lines), which combines the measurement of the current stimulus m_n with the threshold θ_{n-1} from the previous trial, i.e., the reference. Example traces are displayed for intervals of 494 ms (blue) and 847 ms (red). Kernel density estimates are provided for the distributions of the model's stochastic variables (derived from 100 simulation runs for each stimulus in the range $494 - 847$ ms from **Figure 1**). Thick shaded lines in the measurement stage are theoretical distributions. Dotted vertical lines and shaded areas in the reproduction stage give predicted mean ± std.

the present case, we have (cf. Simen et al., 2011):

$$r_T \sim IG\left(\frac{\theta_T}{A_r}, \frac{\theta_T^2}{\sigma_r^2}\right): \qquad \bar{r}_T := \mathrm{E}\,(r \mid T) = \frac{\theta_T}{A_r},$$

$$\mathrm{Var}\,(r_T) := \mathrm{Var}\,(r \mid T) = \frac{\theta_T\,\sigma_r^2}{A_r^3}\,. \quad (5)$$

The reproduced stimulus interval that corresponds to the presentation of a stimulus T is denoted by r_T and θ_T is the threshold in this trial.

2.1.3. Threshold for Reproduction
As already mentioned above, in a trial n the threshold θ_n in the reproduction stage depends on the stimulus' measurement m_n and the threshold θ_{n-1} of the previous trial

$$\theta_n = a\,m_n + (1-a)\,\theta_{n-1},\ \theta_0 = m_0\,. \quad (6)$$

The weight a is limited to the interval $0 < a \le 1$. A value of $a = 0$ has to be excluded since for $a = 0$ only the first stimulus would be taken into account and has an everlasting memory. The formulation in Equation (6) effectively introduces a weighted average preventing unbounded growth. A reference memory is formed and updated on a single trial basis. As we will see later, the memory weight a has an immediate impact on the relation between stimulus and response. The recursive definition in Equation (6) can also be given as an iterative formula

$$\theta_n = a\,m_n + a\sum_{i=1}^{n-1}(1-a)^{n-i}m_i + (1-a)^n m_0\,. \quad (7)$$

2.1.4. Further Conditions for the Model
It is assumed that drifts A_m and A_r are positive numbers. In addition, the drift-diffusion processes are supposed to act in drift-dominated regimes with $A_{m/r}\,T > \sigma_{m/r}\,\sqrt{T}\cdot\mathcal{N}(0,1)$. Otherwise, the measurement stage may yield negative values, resulting in negative thresholds θ, which can not be hit from below. For the sake of simplicity and without loss of generality, the model is not formulated with a lower bound that only allows for positive values. An account of the influence of a lower bound on the first passage time distribution of a drift-diffusion process can be found in Simen et al. (2011).

2.2. Analytical Approximations
Reproduced stimuli in the model are random variables drawn from the distribution of first passage times in the reproduction stage (Equation 5). Determining the distribution of these first passage times $p(r_T)$ is complicated since the threshold θ_T itself is a random variable. Obtaining $p(r_T)$ would thus require calculating $p(r_T) = \int d\theta_T\ p(r_T \mid \theta_T)\,p(\theta_T)$ for which a general solution can not be provided. For a Gaussian threshold distribution the calculations are exemplified by Simen et al. (2011), resulting in smeared-out inverse Gaussian distributions. Qualitatively this results also holds true for other "reasonable" threshold distributions (cf. **Figure 2**). To provide generic analytical solutions for the present model, the section below focuses on expected values and variances.

2.2.1. Expected Value of the Threshold
With randomized stimulus presentations and sufficiently large numbers of preceding trials, we obtain the expected value of the threshold in the current trial from Equation (7)

$$\mathrm{E}\,(\theta_n) = a\,m_n + a\langle m\rangle\sum_{i=1}^{n-1}(1-a)^{n-i} + (1-a)^n m_0\,, \quad (8)$$

with $\langle\cdot\rangle$ denoting the trial average. The sum in Equation (8) is a geometric series and can be rewritten to

$$\mathrm{E}\,(\theta_n) = a\,m_n + a\langle m\rangle\left[\frac{1-(1-a)^n}{1-(1-a)} - 1\right] + (1-a)^n m_0$$

$$= a\,m_n + \langle m\rangle\left[1 - a - (1-a)^n\right] + (1-a)^n m_0\,.$$

We further simplify by taking the limit $n \to \infty$ and get $\bar{\vartheta}_n := \lim_{n\to\infty}\mathrm{E}\,(\theta_n) = a\,m_n + (1-a)\langle m\rangle$. From the last expression we derive the expected value of the threshold in a trial in which the interval T was presented, i.e., $\bar{\vartheta}_T = \mathrm{E}\,(\vartheta \mid \bar{m}_T) = \mathrm{E}\,(\vartheta \mid T)$. Using $\langle m\rangle = A_m\,\langle T\rangle$ and $m_n = \bar{m}_T = A_m\,T$, we end up with

$$\bar{\vartheta}_T = a\,A_m\,T + (1-a)\,A_m\,\langle T\rangle\,. \quad (9)$$

Note, that the average threshold $\bar{\vartheta}_T$ depends on both the current stimulus T and the trial average $\langle T\rangle$. The latter is equal to the mean of the stimulus distribution $\langle T\rangle = \mathrm{E}\,(T)$. The description below therefore uses $\mathrm{E}\,(T)$ instead of $\langle T\rangle$.

2.2.2. Variance of the Threshold
The above calculations only gave the mean threshold for a particular trial. In a next step let us derive from Equation (7) the corresponding variance. Calculating the variance of Equation (7) we obtain a slightly more elaborate geometric series

$$\mathrm{Var}\,(\theta_n) = a^2\,\mathrm{Var}\,(m_n) + a^2\,\mathrm{Var}\,(m)\sum_{j=1}^{n-1}(1-a)^{2j}$$

$$\mathrm{Var}\,(\theta_n) = a^2\,\mathrm{Var}\,(m_n) + a^2\,\mathrm{Var}\,(m)\frac{(1-a)^2 - (1-a)^{2n}}{1-(1-a)^2}\,.$$

Taking the limit $n \to \infty$, yields

$$\mathrm{Var}\,(\vartheta_n) := \lim_{n\to\infty}\mathrm{Var}\,(\theta_n) = a^2\,\mathrm{Var}\,(m_n) + \frac{a(1-a)^2}{2-a}\,\mathrm{Var}\,(m)\,.$$

From the last expression we determine the variance of the threshold in a trial with stimulus interval T, i.e., $\mathrm{Var}\,(\vartheta_T)$. The variance $\mathrm{Var}\,(m_n)$ is given by the variance of the current measurement $\mathrm{Var}\,(m_T) = \sigma_m^2 T$, see Equation (2), and $\mathrm{Var}\,(m)$ is given by Equation (3), i.e., $\mathrm{Var}\,(m) = A_m^2\,\mathrm{Var}\,(T) + \sigma_m^2\,\mathrm{E}\,(T)$. Insertion into the above formula yields

$$\mathrm{Var}\,(\vartheta_T) = a^2\sigma_m^2 T + \left[A_m^2\,\mathrm{Var}\,(T) + \sigma_m^2\,\mathrm{E}\,(T)\right]\frac{a(1-a)^2}{2-a}\,. \quad (10)$$

Thus, similarly to the average threshold (9) its variance also depends on both the current stimulus T and the mean of the stimulus distribution $\mathrm{E}\,(T)$. A third influence comes from the variance of the stimuli $\mathrm{Var}\,(T)$.

2.2.3. Expected Value and Variance for the Reproduction

We can use the solutions for expected value and variance of the threshold ϑ_T from Equations (9, 10) to extend the formulas for the expected value and variance during reproduction in Equation (5). To determine the average reproduced value $\bar{r}_T := \mathrm{E}\,(r \mid T)$ for a stimulus T, we apply the law of total expectation and obtain

$$\bar{r}_T := \mathrm{E}\,(\mathrm{E}\,(r_T \mid \vartheta_T)) = \frac{\bar{\vartheta}_T}{A_r} = a\frac{A_m}{A_r}T + (1-a)\frac{A_m}{A_r}\,\mathrm{E}\,(T) \ . \tag{11}$$

From Equation (11) we also find an expression for the bias corresponding to a stimulus T

$$\mathrm{BIAS}_{r_T} := \bar{r}_T - T = -\left(1 - a\frac{A_m}{A_r}\right)T + (1-a)\frac{A_m}{A_r}\,\mathrm{E}\,(T) \ . \tag{12}$$

Equations (11, 12) directly relate the stimulus T to its reproduced value. Expected value and bias of the reproduced stimuli not only depend on the current stimulus T but also on the expected value of the stimulus distribution $\mathrm{E}\,(T)$. The latter adds an offset to the linear relations. The memory weight a contributes to the slope of the relations and thus determines the strength of the regression effect. Values of a closer to zero result in stronger regression to mean; for values of a closer to one, regression vanishes and reproduction is veridical. As we will see in Section 3, the weight a can be constrained by other model parameters to minimize reproduction errors. Regression and range effects are consequences of such optimization efforts.

Expected value and bias of the reproduction according to Equations (11, 12) also depend on the ratio of drifts from both production and reproduction, A_m and A_r, respectively. Calculating the expectations

$$\mathrm{E}\,(\bar{r}_T) = \frac{A_m}{A_r}\,\mathrm{E}\,(T) \ \text{and} \ \mathrm{E}\,\left(\mathrm{BIAS}_{r_T}\right) = \left(\frac{A_m}{A_r} - 1\right)\mathrm{E}\,(T) \tag{13}$$

shows that for mismatches between the drifts A_m and A_r we get overall deviations between stimuli T and reproductions r_T. These non-zero average biases may explain overall over-estimation (for $A_m/A_r > 1$) and overall under-estimation ($A_m/A_r < 1$), respectively.

To determine the variance $\mathrm{Var}\,(r_T) := \mathrm{Var}\,(r \mid T)$ in a trial in which the stimulus T was presented, we apply the law of total variance and obtain

$$\begin{aligned}\mathrm{Var}\,(r_T) &= \mathrm{E}\,(\mathrm{Var}\,(r_T \mid \vartheta_T)) + \mathrm{Var}\,(\mathrm{E}\,(r_T \mid \vartheta_T)) \\ &= \mathrm{E}\,\left(\frac{\vartheta_T\sigma_r^2}{A_r^3}\right) + \mathrm{Var}\,\left(\frac{\vartheta_T}{A_r}\right) \\ &= \left(\frac{aA_m\sigma_r^2}{A_r^3} + \frac{a^2\sigma_m^2}{A_r^2}\right)T \\ &\quad + (1-a)\,\mathrm{E}\,(T)\frac{A_m\sigma_r^2}{A_r^3} \\ &\quad + \left[A_m^2\,\mathrm{Var}\,(T) + \sigma_m^2\,\mathrm{E}\,(T)\right]\frac{a(1-a)^2}{A_r^2(2-a)} . \end{aligned} \tag{14}$$

Like the variance of the threshold, also the variance $\mathrm{Var}\,(r_T)$ depends on the current stimulus T and the statistics of the stimulus distribution given by $\mathrm{E}\,(T)$ and $\mathrm{Var}\,(T)$. Note that the monotonic relation (14) between stimulus T and variance $\mathrm{Var}\,(r_T)$ of its reproduction is equivalent to scalar variability.

With formulas (11–14), we have a full characterization of the model linking the stimuli T to their reproduced values r_T. The description also details the dependence on the different model parameters, i.e., the internal processing. **Figure 1** gives examples how formulas (11–14) fit to simulations of the model.

3. RESULTS

As displayed in **Figure 1**, the model described in Section 2 can reproduce the typical psychophysical findings for magnitude estimation: regression effect, range effect, scalar variability, and sequential effects. However, it remains open how we can motivate the choice of parameters that fit the psychophysical findings. The upcoming paragraphs focus on this question.

3.1. How to Minimize Reproduction Errors?

Different factors of uncertainty challenge precise magnitude estimation as it is formulated by the model — such as the statistics of the stimuli and internal sources of noise σ_m and σ_r. How could a subject cope with these noise sources to minimize estimation errors?

For optimal magnitude reproduction one needs to minimize the mean squared error between a stimulus T and its reproduced value r_T, i.e., $\mathrm{MSE}_r = \mathrm{E}\left[(r_T - T)^2\right]$. The mean squared error can be partitioned into a variance and a bias term

$$\mathrm{MSE}_r = \mathrm{Var}\,(r) + \mathrm{BIAS}_r^2 \ . \tag{15}$$

The description of the variance $\mathrm{Var}\,(r)$ in Equation (15) depends on the purpose of optimization. In fact, it is not the total variance for the reproduction that should be minimized here. Rather subjects would want to minimize the variability of individual measurements of a particular stimulus $\mathrm{E}\,(\mathrm{Var}\,(r_T)) = \mathrm{E}\,(\mathrm{Var}\,(r \mid T))$; cf. Jazayeri and Shadlen (2010). From Equation (14) the variance $\mathrm{E}\,(\mathrm{Var}\,(r_T))$ is given by

$$\begin{aligned}\mathrm{E}\,(\mathrm{Var}\,(r_T)) &= \left(\frac{A_m\sigma_r^2}{A_r^3} + \frac{a^2\sigma_m^2}{A_r^2}\right)\mathrm{E}\,(T) \\ &\quad + \left[A_m^2\,\mathrm{Var}\,(T) + \sigma_m^2\,\mathrm{E}\,(T)\right]\frac{a(1-a)^2}{A_r^2(2-a)} . \end{aligned} \tag{16}$$

The term BIAS_r^2 in Equation (15) refers to the mean squared or quadratic mean of all biases in a test range, i.e., $\mathrm{E}\left(\mathrm{BIAS}_{r_T}^2\right)$. Using

Equation (12) it is given by

$$BIAS_r^2 = E\left(BIAS_{r_T}^2\right)$$

$$= E\left\{\left[-\left(1-a\frac{A_m}{A_r}\right)T + (1-a)\frac{A_m}{A_r}E(T)\right]^2\right\}$$

$$= \left(1-a\frac{A_m}{A_r}\right)^2 E(T^2) + \left(\frac{A_m^2}{A_r^2}(1+a) - 2\frac{A_m}{A_r}\right)$$

$$(1-a)E(T)^2. \tag{17}$$

With Equations (16, 17) the MSE_r reads as follows

$$MSE_r = E(Var(r \mid T)) + BIAS_r^2$$

$$= \left(\frac{A_m\sigma_r^2}{A_r^3} + \frac{a^2\sigma_m^2}{A_r^2}\right)E(T) + \left[A_m^2 Var(T) + \sigma_m^2 E(T)\right]$$

$$\frac{a(1-a)^2}{A_r^2(2-a)}$$

$$+ \left(1-a\frac{A_m}{A_r}\right)^2 E(T^2) + \left(\frac{A_m^2}{A_r^2}(1+a) - 2\frac{A_m}{A_r}\right)$$

$$(1-a)E(T)^2. \tag{18}$$

Let us explore the possibility that the memory a of the system can be adapted to minimize the mean squared error, i.e., $a_{\min} := \min_a(MSE_r)$. Recall that a is connected to the slope of the relation between stimulus T and its average reproduction $\bar{r}_T$ and thus determines the strength of the regression effect; cf. Equation (11). To find $a_{\min}$ we take the first derivative with respect to a of Equation (18) and set it to zero
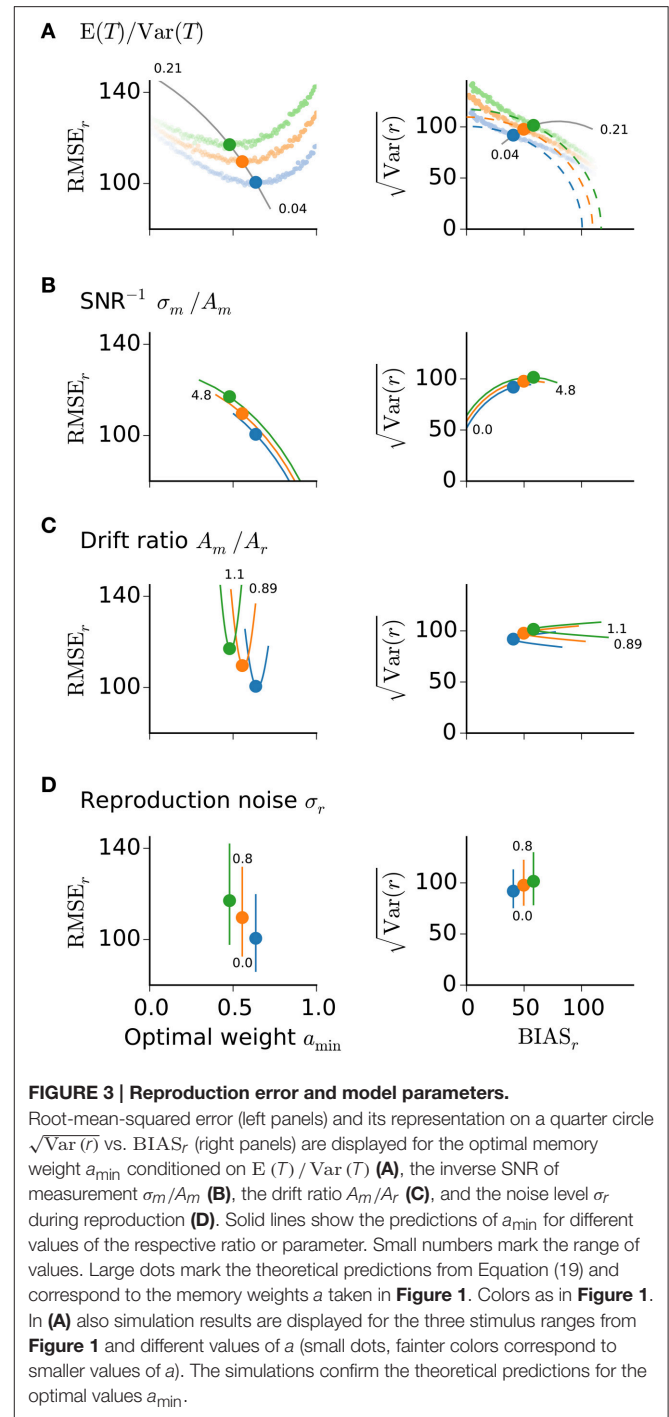
$$0 \overset{!}{=} \frac{d}{da}MSE_r = \frac{d}{da}E(Var(r \mid T)) + \frac{d}{da}BIAS_r^2$$

$$0 = \left(\frac{A_m}{A_r}\right)^2 Var(T)\left[\frac{2}{(2-a)^2} - 2a\right] + \frac{\sigma_m^2}{A_r^2}$$

$$E(T)\frac{2}{(2-a)^2} + 2\left(\frac{A_m}{A_r}\right)^2\left[a - \frac{A_r}{A_m}\right]Var(T)$$

$$0 = \left[Var(T) + \left(\frac{\sigma_m}{A_m}\right)^2 E(T)\right]\frac{1}{(2-a)^2} - \frac{A_r}{A_m}Var(T).$$

Solving for $a = a_{\min}$, we obtain

$$a_{\min} = 2 - \sqrt{\frac{A_m}{A_r}\left[\left(\frac{\sigma_m}{A_m}\right)^2\frac{E(T)}{Var(T)} + 1\right]}. \tag{19}$$

Simulation results confirm the derivation that led to Equation (19); cf. **Figure 3A**.

According to Equation (19) a subject may reduce its overall reproduction error by adjusting the strength of regression, depending on the values of three different relations: (i) the drift ratio A_m/A_r, which may account for overall biases, cf. Equation (13); (ii) the inverse signal-to-noise ratio (SNR) of the measurement σ_m/A_m, quantifying internal noise; and (iii) the inverse of the index of dispersion (variance-to-mean ratio, Fano

FIGURE 3 | Reproduction error and model parameters.
Root-mean-squared error (left panels) and its representation on a quarter circle $\sqrt{Var(r)}$ vs. $BIAS_r$ (right panels) are displayed for the optimal memory weight $a_{\min}$ conditioned on $E(T)/Var(T)$ **(A)**, the inverse SNR of measurement σ_m/A_m **(B)**, the drift ratio A_m/A_r **(C)**, and the noise level σ_r during reproduction **(D)**. Solid lines show the predictions of $a_{\min}$ for different values of the respective ratio or parameter. Small numbers mark the range of values. Large dots mark the theoretical predictions from Equation (19) and correspond to the memory weights a taken in **Figure 1**. Colors as in **Figure 1**. In **(A)** also simulation results are displayed for the three stimulus ranges from **Figure 1** and different values of a (small dots, fainter colors correspond to smaller values of a). The simulations confirm the theoretical predictions for the optimal values $a_{\min}$.

factor) of the stimulus distribution $E(T)/Var(T)$, characterizing the stimulus distribution, and constituting an external source of uncertainty — in contrast to the other two ratios that are due to internal processing. Note that noise in the reproduction, i.e., σ_r, does not influence $a_{\min}$, which intuitively makes sense since the update-step of the memory weight a precedes the reproduction stage.

3.2. Optimality Predicts Range and Regression Effects

To evaluate how the optimal memory weight a_{min} depends on the above ratios, let us consider their individual influences on the reproduction error (**Figure 3**) and determine their interaction (**Figure 4**). **Figure 3** displays the reproduction error as a function of a_{min} for different choices of the model parameters. Instead of the mean squared error its square root $RMSE_r = \sqrt{MSE_r}$ is plotted, which allows for the intuitive visualization of the Pythagorean sum (15) on quarter circles of similar MSE_r-levels (**Figure 3**, right panels).

The MSE_r increases with larger ratio $E(T)/Var(T)$; **Figure 3A**. The dependence serves as an explanation of range effects in magnitude estimation, i.e., dependencies on the stimulus statistics, — an experimentally testable prediction (cf. 4). A larger ratio $E(T)/Var(T)$ corresponds to a narrower stimulus distribution and thus smaller differences between particular stimuli, which in turn are harder to distinguish. This increases uncertainty about the stimuli, which a subject could balance by increasing regression and hence treat different stimuli as more similar (closer to their mean) as they in fact are. Stronger regression is obtained by letting the memory weight a_{min} tend to zero. Note that stronger regression, i.e., smaller a_{min}, results in a stronger change in the $BIAS_r$-component compared to the variance component $\sqrt{Var(r)}$ (**Figure 3A**, right panel). **Figures 4A,B** examines the relation between $E(T)/Var(T)$ and the other model parameters with regard to the optimal weight a_{min}. Only regions with $a_{min} \in (0, 1]$ are displayed to obtain parameter combinations where optimization is possible. The parameter regions where

MSE_r could be optimized shrink with larger $E(T)/Var(T)$ and are further diminished when conditioned on the drift ratio and SNR^{-1} (**Figures 4A,B**).

Larger measurement noise, i.e., $SNR^{-1} = \sigma_m/A_m$, increases MSE_r (**Figure 3B**); to balance this the optimal memory weight a_{min} decreases accordingly (**Figures 3B, 4C,D**). For larger measurement noise, reproduction errors are minimized by increasing regression. The regression effect can thus be interpreted as a strategy to reduce reproduction errors given noisy estimates. In contrast, very precise estimation would lead to veridical judgements about the stimuli. Note, the connection between the inverse SNR and the Weber fraction from psychophysics. Larger SNR^{-1} corresponds to reduced sensory resolution, i.e., lower discriminability, which results in a larger Weber fraction.

The optimal weight a_{min} also depends on the drift ratio A_m/A_r, which if not equal to one, leads to systematic biases, i.e., overall under- or overestimation; cf. Equation (13), and thus larger MSE_r (**Figure 3C**). To compensate for the introduced overall bias (**Figure 3C**, right panel), drift ratios greater than one require smaller a_{min} and drift ratios smaller than one require larger a_{min} (**Figures 4C,D**). Note that the impact of the drift ratio A_m/A_r on a_{min} might point in the opposite direction as that of the external and internal uncertainties, $E(T)/Var(T)$ and σ_m/A_m, respectively.

In summary, the dependence of a_{min} on the noise level during measurement σ_m/A_m predicts the regression effect and the dependence on the stimulus statistics $E(T)/Var(T)$ explains the range effect. The dependence on the ratio of drifts A_m/A_r explains systematic effects like overall over- and underestimation.

FIGURE 4 | Optimality characteristics. (A,B) Optimal weight a_{min} in dependence on the stimulus distribution. **(A)** Optimal weight a_{min} with SNR^{-1} fixed to 4 (*upper panel*) and drift ratio of 1 (*lower panel*), respectively. Same color bar as in **(C**, left). **(B)** Regions of $a_{min} \in (0, 1]$ for three different SNR^{-1} and drift ratios, respectively; thick colored lines $a_{min} = 0$, thin colored lines $a_{min} = 1$. **(C)** Optimal memory weight a_{min} as a function of drift ratio and inverse signal-to-noise ratio. *Left:* Optimal weight a_{min} for the stimulus range $494 - 847$ ms. *Right:* Regions of $a_{min} \in (0, 1]$ for each stimulus range from **Figure 1**. Colors as in **Figure 1**. **(D)** Optimal weight a_{min} as a function of the drift ratio and inverse SNR, respectively, for all three stimulus ranges from **Figure 1**. Again an inverse SNR of 4 and a drift ratio of 1, respectively, have been used. Gray dashed lines mark those values. Simulation data (1000 runs per stimulus) confirm the theoretical prediction (shaded areas of same color, 10% percentile for minimal MSEs).

As already mentioned above noise in the reproduction σ_r does not affect $a_{\min}$; cf. Equation (19). Nevertheless, MSE_r gets increased with larger reproduction noise (**Figure 3D**). Noise in measurement and reproduction therefore differently affects the bias and the variance of stimulus reproduction.

3.3. Explaining Sequential Effects

A fourth class of psychophysical characteristics that was mentioned in the introduction was not considered so far, i.e., effects related to stimulus order (Cross, 1973; Petzschner et al., 2015). Due to the trial-by-trial update rule incorporated in the model, previous trials unavoidably affect the reproduction of the current stimulus. **Figure 1C** exemplifies this via the biases for a particular stimulus conditioned on the stimulus in the previous trial. In general the bias for the current stimulus is proportional to the immediately preceding stimulus. To evaluate this effect analytically, let us reconsider Equation (7). We take out trial $n-1$ from the sum and proceed in similar steps to the derivation in Section "2.2," which lead to Equation (9) and finally to Equation (11). The average response to stimulus T given that stimulus T_{n-1} was presented in the preceding trial is obtained from

$$\bar{r}_{T,T_{n-1}} = a\frac{A_m}{A_r}T + a(1-a)\frac{A_m}{A_r}\Delta T_{n-1} + (1-a)\frac{A_m}{A_r}\,\mathrm{E}\,(T)$$
$$= \bar{r}_T + a(1-a)\frac{A_m}{A_r}\Delta T_{n-1}\,. \tag{20}$$

We express the previous stimulus relative to the mean $\mathrm{E}\,(T)$ here, i.e., $\Delta T_{n-1} = T_{n-1} - \mathrm{E}\,(T)$. The effect of the previous onto the current trial, we evaluate by the corresponding BIAS

$$\mathrm{BIAS}_{T_n,T_{n-1}} = \bar{r}_{T_n,T_{n-1}} - T = \mathrm{BIAS}_{r_T} + a(1-a)\frac{A_m}{A_r}\Delta T_{n-1}\,. \tag{21}$$

Thus, when a stimulus value T_{n-1} larger than $\mathrm{E}\,(T)$ was presented in the previous trial a positive term is added to BIAS_{r_T}. For a stimulus $T_{n-1} < \mathrm{E}\,(T)$ the term is negative and the bias will become smaller (**Figure 1C**).

4. DISCUSSION

The model introduced in the present paper describes magnitude estimation as a two-stage process, measurement and reproduction, consisting of noisy integrators linked by an internal reference (*implicit standard* or *prior*) that is updated on a trial-by-trial basis.

Trial-by-trial update rules have been used by others to explain aspects of magnitude estimation (Hellström, 2003; Dyjas et al., 2012; Bausenhart et al., 2014) and are also at the core of the Bayesian model by Petzschner and Glasauer (2011), where such updating is used to adjust prior knowledge about the stimulus distribution. Iterative updating in the present model estimates the moments of the stimulus distribution to form an internal reference. At least humans are known to be able to maintain (Morgan et al., 2000) and to quickly adapt such an internal reference (Berniker et al., 2010).

Noisy integrative processes well describe decision-making at the behavioral level (Brunton et al., 2013). Moreover, several brain regions show noisy integration during decision-making (Shadlen and Newsome, 2001; Liu and Pleskac, 2011; Shadlen and Kiani, 2013; Hanks et al., 2015) at least at the population level. Whether noisy integration is generated by ramp-like noisy integration in single neurons has been questioned recently (Latimer et al., 2015). In any case, the present model suggests that noisy integration is also crucial to non-binary cognitive demands such as the representation and processing of magnitudes.

4.1. Connection to Psychophysical Effects of Magnitude Estimation

The presented model reproduces the main behavioral characteristics of magnitude estimation **Figure 1**; (Petzschner et al., 2015): Estimates tend toward the mean (regression effect) and this effect scales with the range of stimuli chosen (range effect). Errors monotonically increase with the size of the stimulus (scalar variability). In addition, the sequence in which stimuli are presented influences magnitude judgments. Such sequential effects are by design captured by the model due to the trial-by-trial update of the internal reference. The major insight from this paper therefore is that iterative updating can explain regression and range effects (see also Bausenhart et al., 2014). As such both effects are consequences of strategies to minimize reproduction errors. With larger uncertainty about the stimuli, stronger regression helps to minimize reproduction errors and hence optimizes judgements. Uncertainty may stem from both internal and external sources, whose influence can be evaluated separately by the presented approach.

Internal noise is quantified by the signal-to-noise ratio (SNR) during measurement, i.e., inverse SNR in Equation (19), which corresponds to the Weber fraction in psychophysics and thus the discrimination abilities of a subject. Weber fractions depend on the stimulus modality and are subject-specific. "Modality effects" and individual differences are well known in interval timing literature (Shi et al., 2013). Cicchini et al. (2012) showed that percussionists precisely reproduce temporal intervals and display very weak regression effects in contrast to normal subjects. In addition, the results depended on stimulus modality. For all subject groups, performance was better when intervals were given by auditory rather then visual stimulation. The results of Cicchini et al. (2012) are in line with the present model due to the connection between regression effect and Weber fraction. To explain their experimental data, Cicchini et al. (2012) proposed a Bayesian model that included information about the discrimination abilities (Weber fractions) and obtained very similar results to the present work. Increasing SNR (decreasing Weber fraction) during measurement would require adjusting the drift rate A_m such that it is as large as possible compared to the noise σ_m. However, the drift rate A_m will be limited from above by neuronal and network processes, and related time constants (Murray et al., 2014). Analogous constraints were derived by Cicchini et al. (2012) on the width of the prior distribution. Parkinson patients tested off of their medication display strong regression effects (Malapani et al., 1998, 2002).

In addition, the precision of the responses is reduced. This is in line with the present model, since stronger regression is predicted with reduced precision (i.e., increased variance or reduced SNR).

External uncertainty is due to stimulus context, i.e., the statistics of the stimuli, which is quantified by the ratio between mean and variance of the stimulus distribution in the present model. Larger ratios (narrower stimulus distributions) should lead to stronger regression. Intuitively this means that the width of the stimulus distribution becomes small compared to its mean and individual stimuli can not be discriminated anymore, hence uncertainty increases. The regression effect counteracts this by treating different stimuli similar to their mean. Note the similarity to the Weber-Fechner law, which predicts decreased discriminability with larger stimuli. In line with this view, more difficult magnitude estimation tasks should display stronger regression effects (Teghtsoonian and Teghtsoonian, 1978; Petzschner et al., 2015).

Systematic over- or underestimation are often found in magnitude estimation experiments (for examples see Jazayeri and Shadlen, 2010; Petzschner and Glasauer, 2011; Cicchini et al., 2012). Such differences may, e.g., occur due to attentional and subject-related factors. In the model this would be attributed to differences in the drift rates from measurement and reproduction. Note that only differences are important, absolute scales of (neural) processing (Kiebel et al., 2008; Murray et al., 2014) are not crucial as long as they are similar across processing stages.

The standard deviation is a monotonically increasing function of the stimulus strength in the model presented here; cf. Equation (14) and **Figure 1B**. As such the model is in line with the Weber-Fechner law (scalar variability). However, the Weber-Fechner law predicts a linear increase of variability (standard deviation) as a function of magnitude. According to Equation (14) the increase of the standard deviation is sub-linear (square root) in the present model. This sub-linearity may be rather weak (cf. simulation data and theoretical predictions in **Figure 1B**) and thus may still be in line with experimental data, i.e., differentiating between linearity and weak sub-linearity may be hard from real data. Certain extensions of the model may help to obtain a linear relation. One possibility is introducing a drift ratio A_m/A_r that scales with the stimulus T. Whether scalar variability applies to magnitude estimation without restrictions and across all ranges is not clear. This question is, for example, still a matter of debate in interval timing literature, where non-scalar variability has been reported for specific tasks or situations (like timing while counting or singing; Hinton and Rao, 2004; Grondin and Killeen, 2009).

4.2. Predictions

The formulation of the optimal memory weight $a_{\min}$ according to Equation (19) allows for a number of experimentally testable predictions: (i) Reproduced magnitudes should depend on the stimulus distribution. The experimental studies by Jazayeri and Shadlen (2010), Petzschner and Glasauer (2011), Cicchini et al. (2012) only increased the mean of the stimulus distribution between ranges, which would increase the mean-to-variance ratio and predict stronger regression, i.e., a decrease in a (cf. **Figure 1A**). Stimulus distributions with the same mean but larger variances should result in less regression. Indeed, for their experiments on range effects in loudness and distance estimation, Teghtsoonian and Teghtsoonian (1978) varied the width of the stimulus distribution instead of the mean. They found increasing power exponents with wider stimulus distributions. (ii) Regression to the mean should depend on the discrimination abilities of the individual. Subjects with precise perception of the stimulus magnitude under investigation should show less regression than subjects with reduced abilities; (e.g., Cicchini et al., 2012). This should depend on stimulus modality (Cicchini et al., 2012; Shi et al., 2013) and change with training for a specific task. (iii) Seldom stimuli with a low probability of occurrence and with a magnitude way below or way above the stimulus distribution, should not influence the internal reference. (iv) For strong regression the convergence dynamics of the reference should be much slower then for subjects showing weak regression. The influence of previous stimuli should correlate with the level of regression as well as updating of the references after changing the stimulus distribution within an experimental session.

4.3. Connection to Bayesian Models of Magnitude Estimation

Magnitude estimation has been successfully explained by Bayesian models (Jazayeri and Shadlen, 2010; Petzschner and Glasauer, 2011; Cicchini et al., 2012; Petzschner et al., 2015). The relation between the present work and the Bayesian approaches is not investigated in detail. Nevertheless, some connections shall be discussed. An equivalence between drift-diffusion models and Bayesian frameworks has been described for modeling perceptual decision making (Bitzer et al., 2014) and may also be possible to be established for the model presented here. The measurement phase results in an internal estimate m of a stimulus T drawn from a likelihood distribution $p(m \,|\, T)$. The reproduction process gives a posterior estimate, the reproduced stimulus r, drawn from the distribution $p(r \,|\, m)$. It has to be explored, however, (i) whether the update rule Equation (6) implements a way of connecting both the likelihood $p(m \,|\, T)$ and the posterior $p(r \,|\, m)$ in a Bayes-optimal way; (ii) in how far the update-rules used here and in Petzschner and Glasauer (2011) correspond to each other; and (iii) if the remarkable agreement between the present results and that of the Bayesian description by Cicchini et al. (2012) indicates more than conceptual conformity, i.e., the connection between minimization of reproduction errors and strength of regression, and their modulation by the precision of sensory representations and by the stimulus distribution.

In general, interpreting the regression effect as a means of error minimization shares similarities with concepts like the free-energy principle (Friston, 2010) and information maximization (Linsker, 1990). Error minimization corresponds to the idea of minimizing surprise (free energy) or prediction error and hence maximizing reward.

4.4. Neural Implementation?

Noisy integrative activation patterns are found in several brain regions during decision-making tasks (for a recent review see, e.g., Shadlen and Kiani, 2013). It remains open, however, if such patterns are also present during magnitude estimation as proposed by the model presented here. Neurons sensitive to elapsed time have been shown, for instance, in parietal cortex (Leon and Shadlen, 2003), hippocampus (MacDonald et al., 2011; Sakon et al., 2014), and basal ganglia (Jin et al., 2009; Mello et al., 2015). Neurons in rat hippocampus code for distance covered (Kraus et al., 2013). Single neurons in rat prefrontal cortex show temporally modulated activation patterns during interval timing (Kim et al., 2013; Xu et al., 2014). Such single cell activation patterns may form a set of basis functions to drive noisy integrative processes (c.f. Ludvig et al., 2008; Goldman, 2009; Mello et al., 2015) and may arise in neural networks with balanced excitation and inhibition (Simen et al., 2011), from firing rate adaptation (Reutimann et al., 2004), or from single neuron dynamics (Durstewitz, 2003) — although it has been questioned recently if ramp-like activity is present in single cells (Latimer et al., 2015). It is, furthermore, conceivable to obtain processes akin to noisy integration from state dependent networks (Karmarkar and Buonomano, 2007; Buonomano and Laje, 2010; Laje and Buonomano, 2013). Another question that arises when thinking about a neural implementation of the model introduced in this paper concerns the implementation of the adaptive threshold. It has been suggested from network models of perceptual decision making that adaptive thresholds for noisy integrative processes may be implemented with the help of synaptic plasticity in cortico-striatal circuits (Lo and Wang, 2006; Wei et al., 2015).

5. CONCLUSIONS

The model presented in this paper describes magnitude estimation as two-stages of noisy integration linked by an iteratively updated internal reference memory. Behavioral characteristics well known from magnitude estimation experiments are not only reproduced but also explained as a means of minimizing errors given estimates corrupted by internal and external sources of noise. This paper thus shows that noisy integrative processes may be crucial for cognitive demands beyond perceptual decision making, such as the processing of magnitudes — suggesting an overall computational principle and likely common neural mechanisms that we use to perceive and interpret our environment.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

ACKNOWLEDGMENTS

REFERENCES

Bausenhart, K. M., Dyjas, O., and Ulrich, R. (2014). Temporal reproductions are influenced by an internal reference: explaining the Vierordt effect. *Acta Psychol. (Amst.)* 147, 60–67. doi: 10.1016/j.actpsy.2013.06.011

Berniker, M., Voss, M., and Kording, K. (2010). Learning priors for bayesian computations in the nervous system. *PLoS ONE* 5:e12686. doi: 10.1371/journal.pone.0012686

Bitzer, S., Park, H., Blankenburg, F., and Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Front. Hum. Neurosci.* 81:102. doi: 10.3389/fnhum.2014.00102

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* 113, 700–765. doi: 10.1037/0033-295X.113.4.700

Broderick, T., Wong-Lin, K. F., and Holmes, P. (2009). Closed-form approximations of first-passage distributions for a stochastic decision-making model. *Appl. Math. Res. Express* 2009, 123–141. doi: 10.1093/amrx/\break{}abp008

Brunton, B. W., Botvinick, M. M., and Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science* 340, 95–98. doi: 10.1126/science.1233912

Buonomano, D. V., and Laje, R. (2010). Population clocks: motor timing with neural dynamics. *Trends Cogn. Sci.* 14, 520–527. doi: 10.1016/j.tics.2010.09.002

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

Cross, D. V. (1973). Sequential dependencies and regression in psychophysical judgments. *Percept. Psychophys.* 14, 547–552.

Durstewitz, D. (2003). Self-organizing neural integrator predicts interval times through climbing activity. *J. Neurosci.* 23, 5342–5353.

Dyjas, O., Bausenhart, K. M., and Ulrich, R. (2012). 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. doi: 10.3758/s13414-012-0362-4

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Goldman, M. S. (2009). Memory without feedback in a neural network. *Neuron* 61, 621–634. doi: 10.1016/j.neuron.2008.12.012

Grondin, S., and Killeen, P. R. (2009). Tracking time with song and count: different Weber functions for musicians and nonmusicians. *Atten. Percept. Psychophys.* 71, 1649–1654. doi: 10.3758/APP.71.7.1649

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., and Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* 520, 220–223. doi: 10.1038/nature14066

Hellström, A. (2003). Comparison is not just subtraction: effects of time- and space-order on subjective stimulus difference. *Percept. Psychophys.* 65, 1161–1177. doi: 10.3758/BF03194842

Hinton, S. C., and Rao, S. M. (2004). "One-thousand one... one-thousand two...": chronometric counting violates the scalar property in interval timing. *Psychon. Bull. Rev.* 11, 24–30. doi: 10.3758/BF03206456

Hollingworth, H. L. (1910). The central tendency of judgment. *J. Philos. Psychol. Sci. Methods* 7, 461–469.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55

Jazayeri, M., and Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nat. Neurosci.* 13, 1020–1026. doi: 10.1038/nn.2590

Jin, D. Z., Fujii, N., and Graybiel, A. M. (2009). Neural representation of time in cortico-basal ganglia circuits. *Proc. Natl. Acad. Sci. U.S.A* 106, 19156–19161. doi: 10.1073/pnas.0909881106

Karmarkar, U. R., and Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron* 53, 427–438. doi: 10.1016/j.neuron.2007.01.006

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.\break{}1000209

Kim, J., Ghim, J.-W., Lee, J. H., and Jung, M. W. (2013). Neural correlates of interval timing in rodent prefrontal cortex. *J. Neurosci.* 33, 13834–13847. doi: 10.1523/JNEUROSCI.1443-13.2013

Kraus, B. J., Robinson, R. J., II, White, J. A., Eichenbaum, H., and Hasselmo, M. E. (2013). Hippocampal "time cells": time versus path integration. *Neuron* 78, 1090–1101. doi: 10.1016/j.neuron.2013.04.015

Laje, R., and Buonomano, D. V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* 16, 925–933. doi: 10.1038/nn.3405

Latimer, K. W., Yates, J. L., Meister, M. L. R., Huk, A. C., and Pillow, J. W. (2015). Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science* 349, 184–187. doi: 10.1126/science.aaa4056

Leon, M. I., and Shadlen, M. N. (2003). Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* 38, 317–327. doi: 10.1016/S0896-6273(03)00185-5

Linsker, R. (1990). Perceptual neural organization: some approaches based on network models and information theory. *Annu. Rev. Neurosci.* 13, 257–281. doi: 10.1146/annurev.ne.13.030190.001353

Liu, T., and Pleskac, T. J. (2011). Neural correlates of evidence accumulation in a perceptual decision task. *J. Neurophysiol.* 106, 2383–2398. doi: 10.1152/jn.00413.2011

Lo, C.-C., and Wang, X.-J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat. Neurosci.* 9, 956–963. doi: 10.1038/nn1722

Ludvig, E. A., Sutton, R. S., and Kehoe, E. J. (2008). Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. *Neural Comput.* 20, 3034–3054. doi: 10.1162/neco.2008.11-07-654

MacDonald, C. J., Lepage, K. Q., Eden, U. T., and Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron* 71, 737–749. doi: 10.1016/j.neuron.2011.07.012

Malapani, C., Deweer, B., and Gibbon, J. (2002). Separating storage from retrieval dysfunction of temporal memory in Parkinson's disease. *J. Cogn. Neurosci.* 14, 311–322. doi: 10.1162/089892902317236920

Malapani, C., Rakitin, B., Levy, R. S., Meck, W. H., Deweer, B., Dubois, B., et al. (1998). Coupled temporal memories in Parkinson's disease: a dopamine-related dysfunction. *J. Cogn. Neurosci.* 10, 316–331.

Mello, G. B. M., Soares, S., and Paton, J. J. (2015). A scalable population code for time in the striatum. *Curr. Biol.* 25, 1113–1122. doi: 10.1016/j.cub.2015.02.036

Merchant, H., Harrington, D. L., and Meck, W. H. (2013). Neural basis of the perception and estimation of time. *Annu. Rev. Neurosci.* 36, 313–336. doi: 10.1146/annurev-neuro-062012-170349

Morgan, M. J., Watamaniuk, S. N., and McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Res.* 40, 2341–2349. doi: 10.1016/S0042-6989(00)00093-6

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., et al. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663. doi: 10.1038/nn.3862

Petzschner, F. H., and Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *J. Neurosci.* 31, 17220–17229. doi: 10.1523/JNEUROSCI.2028-11.2011

Petzschner, F. H., Glasauer, S., and Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends Cogn. Sci.* 19, 285–293. doi: 10.1016/j.tics.2015.03.002

Reutimann, J., Yakovlev, V., Fusi, S., and Senn, W. (2004). Climbing neuronal activity as an event-based cortical representation of time. *J. Neurosci.* 24, 3295–3303. doi: 10.1523/JNEUROSCI.4098-03.2004

Sakon, J. J., Naya, Y., Wirth, S., and Suzuki, W. A. (2014). Context-dependent incremental timing cells in the primate hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18351–18356. doi: 10.1073/pnas.1417827111

Seabold, S., and Perktold, J. (2010). "Statsmodels: econometric and statistical modeling with python," in *Proceedings of the 9th Python in Science Conference*, eds S. van der Walt and J. Millman, 57–61. Available online at: http://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf

Shadlen, M. N., and Kiani, R. (2013). Decision making as a window on cognition. *Neuron* 80, 791–806. doi: 10.1016/j.neuron.2013.10.047

Shadlen, M. N., and Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* 86, 1916–1936. doi: 10.3410/f.1001494.23207

Shi, Z., Church, R. M., and Meck, W. H. (2013). Bayesian optimization of time perception. *Trends Cogn. Sci.* 17, 556–564. doi: 10.1016/j.tics.2013.09.009

Simen, P., Balci, F., de Souza, L., Cohen, J. D., and Holmes, P. (2011). A model of interval timing by neural integration. *J. Neurosci.* 31, 9238–9253. doi: 10.1523/JNEUROSCI.3121-10.2011

Teghtsoonian, R., and Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Percept. Psychophys.* 24, 305–314.

Tuckwell, H. C. (1988). *Introduction to Theoretical Neurobiology: Volume 2, Nonlinear and Stochastic Theories.* Cambridge, MA: Cambridge University Press.

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

Weber, E. H. (1851). "Die lehre vom tastsinne und gemeingefühle auf versuche gegründet," in *Handwörterbuch der Physiologie*, ed R. Wagner (Braunschweig: Vieweg), 481–588.

Wei, W., Rubin, J. E., and Wang, X.-J. (2015). Role of the indirect pathway of the basal ganglia in perceptual decision making. *J. Neurosci.* 35, 4052–4064. doi: 10.1523/JNEUROSCI.3611-14.2015

Xu, M., Zhang, S.-y., Dan, Y., and Poo, M.-M. (2014). Representation of interval timing by temporally scalable firing patterns in rat prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 480–485. doi: 10.1073/pnas.1321314111