NB}
DT

ORIGINAL ARTICLE

Journal Section

Performance of normative and approximate evidence accumulation on the dynamic clicks task

Adrian E. Radillo[1],* | **Alan Veliz-Cuba**[2],* |
Krešimir Josić[3,4],† | **Zachary P. Kilpatrick**[5,6],†

[1]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104

[2]Department of Mathematics, University of Dayton, Dayton, OH 45469

[3]Departments of Mathematics and Biology and Biochemistry, University of Houston, Houston, TX 77204

[4]Department of BioSciences, Rice University, Houston, TX 77251, USA

[5]Department of Applied Mathematics, University of Colorado, Boulder, CO 80309

[6]Department of Physiology and Biophysics, University of Colorado School of Medicine, Aurora, CO 80045

Correspondence

Zachary P. Kilpatrick, Department of Applied Mathematics, University of Colorado, Boulder, CO 80309
Email: zpkilpat@colorado.edu

The aim of a number of psychophysics tasks is to uncover how mammals make decisions in a world that is in flux. Here we examine the characteristics of ideal and near–ideal observers in a task of this type. We ask when and how performance depends on task parameters and design, and, in turn, what observer performance tells us about their decision-making process. In the dynamic clicks task subjects hear two streams (left and right) of Poisson clicks with different rates. Subjects are rewarded when they correctly identify the side with the higher rate, as this side switches unpredictably. We show that a reduced set of task parameters defines regions in parameter space in which optimal, but not near-optimal observers, maintain constant response accuracy. We also show that for a range of task parameters an approximate normative model must be finely tuned to reach near-optimal performance, illustrating a potential way to distinguish between normative models and their approximations. In addition, we show that using the negative log-likelihood and the 0/1-loss functions to fit these types of models is not equiva-

---

*co-first authors

†co-last authors

Code Availability: Codes developed to produce figures are available at `https://github.com/aernesto/NBDT_dynamic_clicks`

lent: the 0/1-loss leads to a bias in parameter recovery that increases with sensory noise. These findings suggest ways to tease apart models that are hard to distinguish when tuned exactly, and point to general pitfalls in experimental design, model fitting, and interpretation of the resulting data.

KEYWORDS
decision-making, Poisson clicks, Bayesian inference, dynamic environment, model identifiability

1 | INTRODUCTION

Decision-making tasks are widely used to probe the neural computations that underlie behavior and cognition (Luce, 1986; Gold and Shadlen, 2007). Mathematical models of optimal decision-making (normative models)[1] have been key in helping us understand tasks that require the accumulation of noisy evidence (Wald and Wolfowitz, 1948; Gold and Shadlen, 2002; Bogacz et al., 2006). Such models assume that an observer integrates a sequence of noisy measurements to determine the probability that one of several options is correct (Wald and Wolfowitz, 1948; Beck et al., 2008; Veliz-Cuba et al., 2016).

The random dot motion discrimination (RDMD) task is a prominent example, in which the neural substrates of the evidence accumulation process can be identified in cortical recordings (Ball and Sekuler, 1982; Britten et al., 1992; Roitman and Shadlen, 2002). The associated normative models take the form of tractable stochastic differential equations (Ratcliff, 1978; Bogacz et al., 2006), and have been used to explain behavioral data (Ratcliff and McKoon, 2008; Krajbich and Rangel, 2011). Neural correlates of subjects' decision processes display striking similarities with these models (Shadlen and Newsome, 1996; Huk and Shadlen, 2005), although a clear link between the two is still under debate (Latimer et al., 2015; Shadlen et al., 2016).

Poisson clicks tasks (Brunton et al., 2013; Odoemene et al., 2018) have recently become popular in studying the cortical computations underpinning mammalian perceptual decision-making. Neural activity during such tasks also appears to reflect an underlying evidence accumulation process (Hanks et al., 2015). The corresponding normative models and their approximations are low-dimensional and computationally tractable. This makes the task well-suited to the analysis of data in high-throughput experiments (Brunton et al., 2013). Piet et al. (2018) have extended the clicks task to a dynamic environment to understand how animals adjust their evidence accumulation strategies when older evidence decreases in relevance. Glaze et al. (2015) carried out a similar study in an extension of the RDMD task. Both studies concluded that subjects are capable of implementing evidence accumulation strategies that adapt to the timescale of the environment.

However, identifying the specific strategy subjects use to solve a decision task can be difficult because different strategies can lead to similar observed outcomes (Ratcliff and McKoon, 2008). How to set task parameters to best identify a subject's evidence accumulation strategy has not been studied systematically, especially in dynamic environments (Ratcliff et al., 2016). Here, we focus on the dynamic clicks task and aim to understand what task parameters (or combinations of parameters) determine performance, and under what conditions different strategies can be identified.

In the dynamic clicks task, two streams of auditory clicks are presented simultaneously to a subject, one stream per

---

[1] We will use the phrases 'optimal model,' 'optimal observer,' 'normative model,' and 'ideal observer' interchangeably, as they refer to the best possible model for a given set of task and observation constraints.

ear (Piet et al., 2018; Brunton et al., 2013). Each click train is generated by a Markov-modulated Poisson process (Fischer and Meier-Hellstern, 1992) whose click arrival rates switch between two possible values (λ_{low} vs. λ_{high}). The two streams have distinct rates which switch at discrete points in time according to a memoryless process with hazard rate h. Thus streams played in different ears always have distinct rates ($\lambda^L(t) \neq \lambda^R(t)$). The subject must choose the stream with highest instantaneous rate when interrogated at a time T, which ends the trial. Switches occur at random times that are not signaled to the subjects, who must therefore base their decision on the observed sequences of Poisson clicks alone. The rate h at which the environment changes is a latent variable that needs to be learned for optimal performance. In this study, however, our observer models always use a constant rate of change for their environment [2].

We analyze the normative model of the dynamic clicks task to shed light on how its response accuracy depends on task parameters, as this is a measure commonly used when fitting to behavioral data (Piet et al., 2018). As shown in Section 2, the ideal observer accumulates evidence from each click to update their log likelihood ratio (LLR) of the two choices. Each click corresponds to a pulsatile increase or decrease in the LLR. Importantly, evidence must be discounted at a rate that accounts for the timescale of environmental changes.

The main goal of this work is to identify how task parameters shape an ideal observer's response accuracy, and the identifiability of evidence accumulation models. We find effective parameters that can be fixed to keep the accuracy of the ideal observer constant.[3] One such parameter is the signal-to-noise ratio (SNR) of the clicks during a single epoch between changes and the other is hT, the trial length T rescaled by the hazard rate h. These two parameters fully determine the accuracy of an optimal observer interrogated at the end of the trial (Section 3), as well as response accuracy conditioned on the time since the final change point of a trial (Section 4).

While the normative model determines the optimal strategy, subjects may also use heuristics or approximations that are potentially simpler to implement. The accuracy of approximate models may also be more sensitive to parameter changes, so fitting procedures converge more rapidly. As an example we consider a linear model, which has been previously fit to data from subjects performing dynamic decision tasks (Piet et al., 2018; Glaze et al., 2015), and has also been studied as an approximation to normative evidence accumulation (Veliz-Cuba et al., 2016). To obtain response accuracy close to that of the normative model, the discounting rate of the linear model needs to be tuned for different click rates and hazard rates (Section 5). In contrast, the discounting rate in the normative (nonlinear) model equals the hazard rate. Moreover, the linear model's accuracy is more sensitive to changes in its evidence-discounting parameter than the nonlinear model. [4] This effect is most pronounced at intermediate SNR values, suggesting a task parameter range where the two models can be distinguished.

Lastly, we ask how model parameters can be inferred from subject responses. Using maximum likelihood fits of the models to choice data, we show that the fit discounting parameters are closer to the true parameter in the linear model compared to the nonlinear model (Section 6). This is expected, since the response accuracy of the nonlinear model depends weakly on its discounting parameter. We also explore the impact of the loss function on model fitting, and show that in the presence of sensory noise using a 0/1-loss function results in a systematic bias in parameter recovery (Section 7). The 0/1-loss function gives a one unit penalty on trials in which the decision predicted by the model and the data disagree, and no penalty when they agree. Therefore, minimizing this loss function leads to models that best match the trial-to-trial responses in the data rather than the response accuracy.

Ultimately, our findings point to ways of identifying task parameters for which subjects' decision accuracy is

---

[2] See Radillo et al. (2017) for an optimal observer that can learn the hazard rate h in a dynamic version of the RDMD task. This approach can be extended to the case of the dynamic clicks tasks as in Radillo (2018).

[3] We use the term "accuracy" to refer to the percentage of correct choices for a given model and parameter set. This is our primary measure of a model's performance on the task.

[4] The 'nonlinear' model here refers to the family of models obtained by tuning the discounting rate away from the value defining the normative model. This detuning results in a model that is not normative.

sensitive to the mode of evidence accumulation they use in fluctuating environments. We also show how using different models and different data fitting methods can lead to divergent results, especially in the presence of sensory noise. We argue that similar issues can arise whenever we try to interpret data from decision-making tasks.

2 | NORMATIVE MODEL FOR THE DYNAMIC CLICKS TASK

In the dynamic clicks task an observer is presented with two Poisson click streams, $s^L(t)$ and $s^R(t)$ ($0 < t \leq T$), and needs to decide which of the two has a higher rate (Brunton et al., 2013). The rates of the two streams are not constant, but change according to a hidden, continuous-time Markov chain, $x(t)$, with binary state space $\{x^R, x^L\}$. The frequency of the switches is determined by the *hazard rate*, h, so that $P(x(t + dt) \neq x(t)) = h \cdot dt + o(dt)$. The left and right rates, $\lambda^L(t)$ and $\lambda^R(t)$, can each take on one of two values, $\{\lambda_{\text{high}}, \lambda_{\text{low}}\}$ with $\lambda_{\text{high}} > \lambda_{\text{low}} > 0$. When $x(t) = x^L$, we have $(\lambda^L(t), \lambda^R(t)) = (\lambda_{\text{high}}, \lambda_{\text{low}})$, and when $x(t) = x^R$ the opposite is true. Therefore $x(t) = x^k$ means that stream k has the higher rate at time t: $\lambda^k(t) = \lambda_{\text{high}}$ (Fig. 1A). The observer is prompted to identify the side of the higher rate stream, $x(T)$, at a random time T. The interrogation time, T, is sampled ahead of time by the experimenter for each trial and is unknown to the subject. We refer the reader to Piet et al. (2018) and Brunton et al. (2013) for more details about the experimental setup.

This task is closely related to the *filtering of a Hidden Markov Model* studied in the signal processing literature (Cappé et al., 2005; Rabiner and Juang, 1986). For a single, 2-state Markov-modulated Poisson process (Fischer and Meier-Hellstern, 1992), the filtering problem was solved by Rudemo (1972) – see also (Snyder, 1975) for review and extensions. This filtering problem corresponds to a task in which a single, variable rate click stream is presented to the observer who has to report whether at some time T the rate is high or low. In the present case, the observer is presented with two coupled Markov-modulated Poisson processes. The normative model reduces to that considered by Rudemo (1972) when we consider a single stream version of the task, so our approach can be considered a generalization.

Assuming the Poisson rates $\{\lambda_{\text{high}}, \lambda_{\text{low}}\}$, and the hazard rate, h, are known, a normative model for the inference of the hidden state, $x(t)$, has been derived by Piet et al. (2018). The resulting model can be expressed as an ordinary differential equation (ODE) describing the evolution of the LLR of the two environmental states:

$$y_t := \log \frac{P(x(t) = x^R | s^R(t), s^L(t))}{P(x(t) = x^L | s^R(t), s^L(t))}. \tag{1}$$

For completeness, we present the derivation in Appendix A, yielding the same ODE as Piet et al. (2018):

$$\frac{dy_t}{dt} = \kappa \underbrace{\left[\sum_{i=1}^{\infty} \delta(t - t_i^R) - \sum_{j=1}^{\infty} \delta(t - t_j^L) \right]}_{\text{right and left click streams}} - \underbrace{2h \sinh(y_t)}_{\text{nonlinear discounting}}, \tag{2}$$

where $\kappa := \log(\lambda_{\text{high}}/\lambda_{\text{low}})$ is the evidence gained from a click, $\delta(t)$ is the Dirac delta function centered at 0, and t_i^R (resp. t_j^L) is the i-th right click (resp. j-th left click).

Eq. (2) has an intuitive interpretation: A click provides evidence that the higher rate stream is on the side at which the click was heard. Thus, a click heard on the right (left) results in a positive (negative) increment in the LLR (Fig. 1B). Since the environment is volatile, as evidence recedes into the past it becomes less relevant. In Eq. (2) each click is followed by a superlinear decay to zero. Note that the discounting term only depends on the current LLR, y_t, and the hazard rate, h, and not on the click rates.
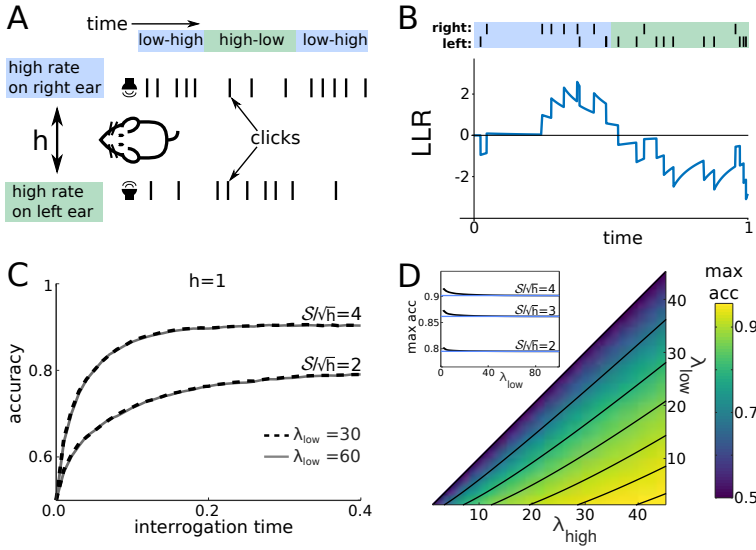
FIGURE 1 **A:** Schematic of the dynamic clicks task from Piet et al. (2018). **B:** A single trajectory of the log-likelihood ratio (LLR), y_t, during a trial. The click streams and environment state are shown above the trajectory. **C:** Response accuracy of the ideal observer as a function of interrogation time for two distinct SNR values, $\mathcal{S}/\sqrt{h}$, defined in Eq. (6). Two distinct pairs of click rates used in simulations ($\lambda_{low} = 30$ and 60 Hz) were chosen to match each SNR at hazard rate $h = 1$ Hz, resulting in overlaying dashed ($\lambda_{low} = 30$Hz) and solid ($\lambda_{low} = 60$Hz) lines. For $\mathcal{S}/\sqrt{h} = 2$, we take $(\lambda_{high}, \lambda_{low}) = (58.17, 30)$ and $(97.67, 60)$; for $\mathcal{S}/\sqrt{h} = 4$, $(\lambda_{high}, \lambda_{low}) = (70, 30)$ and $(112.54, 60)$. Fixing c_1 and c_2 in Eq. (6) yields a match in accuracy for both pairs of click rates. As time evolves during the trial, accuracy saturates. Note hT and $\mathcal{S}/\sqrt{h}$ jointly determine accuracy. **D:** Maximal accuracy of the ideal observer at $T \gg 1$ is constant along level curves (black lines) of $\mathcal{S}/\sqrt{h}$ (seen as a function of $(\lambda_{low}, \lambda_{high})$ with constant h) across a wide range of parameters, consistent with Eq. (6). We only show the $(\lambda_{low}, \lambda_{high})$ region where $0 < \lambda_{low} < \lambda_{high}$. Level curves of $\mathcal{S}$ are oblique parabolas in the $(\lambda_{low}, \lambda_{high})$ plane. **D(Inset):** Level curves slightly underestimate accuracy for small λ_{high} and λ_{low} (black: maximal accuracy; blue: maximal accuracy for large values of λ_{high} and λ_{low}). See Appendix F for details on Monte Carlo simulations.

Performance on the task may increase with the informativeness of each click, $\kappa = \log(\lambda_{high}/\lambda_{low})$. However, κ alone does not predict the response accuracy (i.e. the fraction of correct trials) of the normative model (Brunton et al., 2013; Piet et al., 2018). In the next section, we will show that an ideal observer's response accuracy is determined by the click frequencies ($\lambda_{high}, \lambda_{low}$) and the hazard rate h: A sequence of a few very informative clicks may provide as much evidence as many clicks, each carrying little information. But if the environmental hazard rate is high, even informative clicks quickly lose their relevance.

The LLR, y_t, contains all the information an ideal observer has about the present state of the environment, given the observations (Gold and Shadlen, 2002). If interrogated at time $t = T$, sign(y_T) = ±1, determines the most likely current state (x^R for +1 and x^L for −1), and therefore the response of an optimal decision maker. In the following, we will show that two effective parameters govern the response accuracy of the optimal observer.

3 | THE SIGNAL-TO-NOISE RATIO OF DYNAMIC CLICKS

Four parameters characterize the dynamic clicks task: the hazard rate, h, duration of a trial, *i.e.* interrogation time, T, the low click rate, λ_{low}, and the high click rate, λ_{high}. However, we next show that only two effective parameters typically govern an ideal observer's performance (Fig. 1C,D): the product of the interrogation time and the hazard rate, hT, and the signal-to-noise ratio (SNR) of the dynamic stimulus. The former corresponds to the mean number of switches in a trial, and the latter combines the click rates λ_{low} and λ_{high} into a Skellam–type SNR (Eq. (4) below), scaled by the hazard rate h (Eq. (6)).

To motivate our definition, consider first the case of a static environment, $h = 0$ Hz, for which the normative model is given by Eq. (2) without the nonlinear term. Since κ does not affect the sign of y_t, response accuracy depends entirely on the difference in click counts $N^R(t) - N^L(t)$, where $N^j(t)$ are the counting processes associated with each click stream. Thus we can define the difference in click counts as the *signal*, and the SNR as the ratio between the signal mean and standard deviation at time T (Skellam, 1946),

$$\text{SNR}_0^T := \frac{\text{E}[N^R(T)] - \text{E}[N^L(T)]}{\sqrt{\text{Var}[N^R(T)] + \text{Var}[N^L(T)]}} = \frac{T\lambda_{\text{high}} - T\lambda_{\text{low}}}{\sqrt{T\lambda_{\text{high}} + T\lambda_{\text{low}}}} = \mathcal{S} \cdot \sqrt{T}, \tag{3}$$

where

$$\mathcal{S} := \frac{\lambda_{\text{high}} - \lambda_{\text{low}}}{\sqrt{\lambda_{\text{high}} + \lambda_{\text{low}}}}. \tag{4}$$

In a dynamic environment, the volatility of the environment, governed by the hazard rate, h, also affects response accuracy. The environment can switch states immediately before the interrogation time, T, decreasing response accuracy. This suggests that accuracy will not only be determined by the click rates, but also by the length of time the environment remains in the same state prior to interrogation. Using this fact and the definition of SNR in a static environment, we determine the statistics for the difference in the number of clicks between the high- and low-rate streams during the final epoch preceding interrogation (for derivation details see Appendix B). Averaging over the Poisson distributions characterizing the click numbers, and the epoch length distribution yields a nonlinear expression representing the SNR that involves $\mathcal{S}/\sqrt{h}$, and the rescaled trial time hT:

$$F(hT, \mathcal{S}/\sqrt{h}) = \frac{(1 - e^{-hT})\mathcal{S}/\sqrt{h}}{\sqrt{(1 - 2hTe^{-hT} - e^{-2hT})\mathcal{S}^2/h + 1 - e^{-hT}}}. \tag{5}$$

The unitless measure of trial duration, hT, characterizes the timescale of the evolution of the LLR, y_t. As accuracy should not depend on the units in which we measure time, this is a natural measure of the evidence accumulation period [5]. As indicated, $F(hT, \mathcal{S}/\sqrt{h})$ only depends on hT and $\mathcal{S}/\sqrt{h}$. We therefore predict that optimal observer response accuracy is determined by the following two parameter combinations,

$$\mathcal{S}/\sqrt{h} = \frac{\lambda_{\text{high}} - \lambda_{\text{low}}}{\sqrt{h(\lambda_{\text{high}} + \lambda_{\text{low}})}} = c_1 \qquad \text{and} \qquad hT = c_2. \tag{6}$$

Henceforth, we will refer to $\mathcal{S}/\sqrt{h}$ as the SNR and hT as rescaled trial time. Note that the term $\mathcal{S}/\sqrt{h}$ can also be realized as a SNR of Eq. (2) by performing a diffusion approximation, and computing the SNR of the corresponding drift-diffusion

---

[5] This is related to dimensional analysis often used when studying physical models (Langhaar, 1980).

signal (See Appendix C).

Fig. 1C shows examples in which the ideal observer's response accuracy is constant when SNR and hT are fixed. Accuracy is computed as the fraction of trials at which the observer's belief, y_t, matches the underlying state, $x(t)$, at the interrogation time, T, that is the fraction of trials for which $\text{sign}(y_T) = x(T)$. The accuracy as a function of T and $h = 1$ remains constant if we change λ_{high} and λ_{low}, but keep $\mathcal{S}$ fixed. As the interrogation time T is increased, the accuracy saturates to a value below 1 (Fig. 1C), consistent with previous modeling studies of decision-making in dynamic environments (Glaze et al., 2015; Veliz-Cuba et al., 2016; Radillo et al., 2017; Piet et al., 2018). Evidence discounting limits the magnitude of the LLR, y_t. Hence a sequence of low rate clicks can lead to errors, especially for low SNR values. Moreover, on some trials the state, $x(t)$, switches close to the interrogation time T. As it may take multiple clicks for y_t to change sign after a change point (See Fig. 1B), this can also lead to an incorrect response.

In Fig. 1D we show that the maximal accuracy (obtained for T sufficiently large) as a function of λ_{high} and λ_{low} (colormap), is approximately constant along SNR level sets (black oblique curves). This correspondence is not exact when λ_{high} and λ_{low} are small (Fig. 1D inset), and we conjecture that this is because higher order statistics of the signal determine response accuracy in this parameter range. As discussed in Appendix C, for large λ_{high} and λ_{low} we can use a diffusion approximation for the dynamics of Eq. (2). When λ_{high} and λ_{low} are small, the diffusion approximation does not apply, and response accuracy is characterized by features of the signal beyond its mean and variance. Since the SNR only describes the ratio between the mean and standard deviation of the stimulus, it cannot capture the impact of higher order statistics on accuracy at low click rates. Nonetheless, the SNR predicts response accuracy well.

The consequences of these observations are twofold: Two parameter combinations determine optimal performance, potentially simplifying experiment design. To ensure coverage of different response accuracy regimes, we can initially vary SNR and hT. To increase the accuracy of an ideal observer, it is not sufficient to increase both click rates, for instance, since the SNR stays constant if λ_{high} and λ_{low} follow the parabolas shown in Fig. 1D. Second, this approach makes testable predictions about the accuracy of an optimal observer: If we change parameters so that SNR and hT are fixed, and a subject's accuracy is affected, this indicates that the subject may not have learned the hazard rate, h or is using a suboptimal discounting model.

4 | POST CHANGE-POINT DECISIONS DEPEND ON SNR

To understand how an optimal observer adapts to environmental changes, we next ask how their fraction of correct responses depends on the final time, T_{fin}, between the last change point preceding a decision and the decision itself (Fig. 2A). Overall accuracy again depends on both SNR and rescaled trial time hT. In addition, for sufficiently long trials, accuracy as a function of time since the last change point depends only on the rescaled time since the change point, hT_{fin} and the SNR.

If the click rates, λ_{high} and λ_{low}, are varied, but $\mathcal{S}$ and h are held fixed, the accuracy as a function of T_{fin} remains unchanged (Fig. 2B, for $h = 1$, $\mathcal{S} = 2$). On the other hand, accuracy changes if we fix $\mathcal{S}/\sqrt{h}$ (SNR) but vary h (Fig. 2B, left inset). With fixed $\mathcal{S}/\sqrt{h}$, accuracy depends only on the rescaled time since the last change point, hT_{fin} (Fig. 2B, right inset). Thus, while absolute accuracy depends on the total length of the trial, T, measured in units of average epoch length, $1/h$, accuracy relative to the last change point depends only on the elapsed time, T_{fin}, measured in the same units.

Glaze et al. (2015) introduced the notion of an accuracy *crossover effect* in the dynamic RDMD task: The normative model predicts that after a change point observers update their belief more slowly, but eventually reach higher accuracy at low compared to high hazard rates. Thus plotting the maximal accuracy against time since the last change point for different hazard rates results in curves that cross. Behavioral data indicates that human observers behave according to
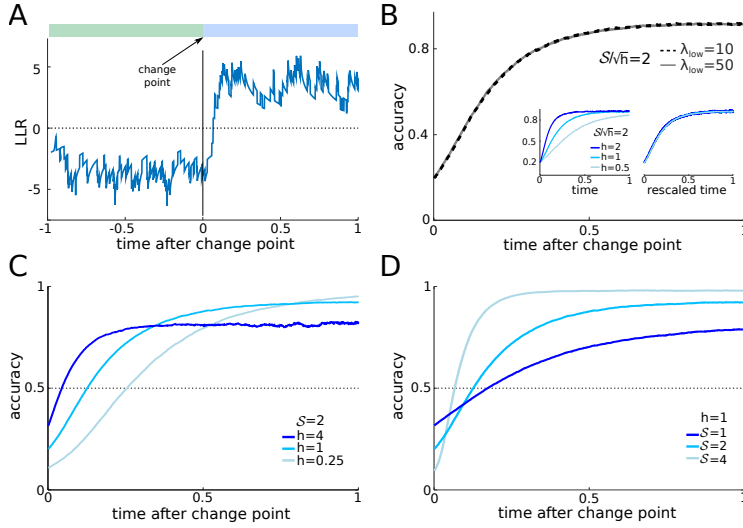
FIGURE 2 **A:** After a change point, evidence in favor of the correct hypothesis has to be accumulated before optimal observers change their belief. Thus, there is a delay between the change point and the sign change of the LLR. **B:** The accuracy as a function of time following a change point is the same when $S/\sqrt{h}$ and h are fixed ($h = 1$ in the plot). **B(inset):** Furthermore accuracy is constant for fixed $S/\sqrt{h}$ and rescaled response time hT_{fin} (left: time without rescaling, right: time after rescaling). **C:** For lower hazard rates, h, accuracy at change point time is lower, and requires a longer time to reach 0.5. **D:** As the strength of evidence, S, increases, accuracy at the time of the change point is lower, but increases more rapidly. In panels C and D the value of $S/\sqrt{h}$ is the same for curves of the same color. Rescaling the timescale of curves in panel C by h would yield the curves in panel D.

this prediction (Glaze et al., 2015, 2018).

A similar crossover effect also occurs in the dynamic clicks task: Accuracy just after a change point is lower for small hazard rates, h, (Fig. 2C) and takes longer to reach 50%, but saturates at a higher level compared to more volatile environments. In slow environments, the optimal observer integrates evidence over a longer timescale ($1/h$), leading to more reliable estimates of the state, $x(t)$. But this increased certainty comes at a price, as it requires more time to change the observer's belief after a change point. Similarly, in environments with stronger evidence (larger S, Eq. (4)), accuracy immediately following a change point is lower, since state estimates, and hence the beliefs are more reliable compared to trials with weak evidence (Fig. 2D). However, stronger evidence also causes a rapid increase in accuracy, which then saturates at a higher level than on trials with weaker evidence (lower S). Therefore, both evidence quality, and environmental volatility determine accuracy after a change point.

We conclude that accuracy after a change point is characterized by SNR ($S/\sqrt{h}$) and the rescaled time since the change point, hT_{fin}. This only holds when trials are sufficiently long, and the belief at trial outset does not affect accuracy. In addition, increasing SNR lowers accuracy immediately after a change point, and increases the recovery of accuracy to a higher saturation point (Fig. 2D). On the other hand decreasing volatility, h, while fixing S (Fig. 2C) leads to lower accuracy immediately after a change point, and higher saturation. However, the rate at which accuracy is recovered after a change point *decreases* with decreasing h.

These are again characteristics of an optimal observer, and deviations from these predictions indicate departures from optimality.
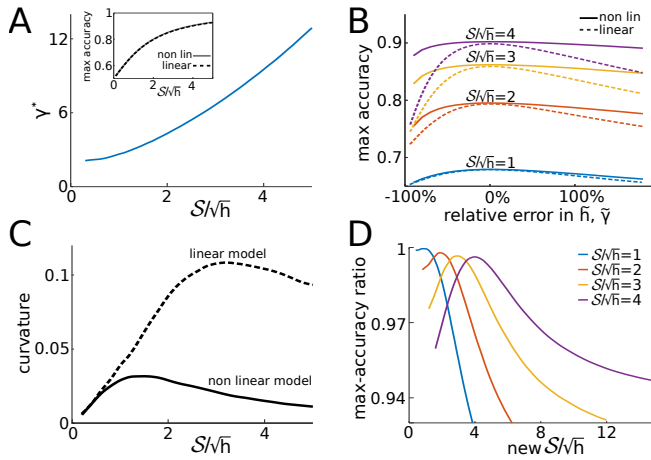
FIGURE 3 **A:** Optimal linear discounting rate, γ^* in Eq. (7), as a function of $\mathcal{S}/\sqrt{h}$. **A(inset):** The accuracy of the linear and nonlinear model are nearly identical over a wide range of SNR values, $\mathcal{S}/\sqrt{h}$, when the linear discounting rate is set to γ^* (here $h = 1$ is fixed). **B:** Response accuracy near the optimal discounting rate for the linear model (dashed), and assumed hazard rates for the nonlinear model (solid) for several SNR values ($h = 1$). The linear model is more sensitive to relative changes in the discounting rate. The relative error is defined as $100(\tilde{h} - h)/h$ for the nonlinear model and $100(\tilde{\gamma} - \gamma^*)/\gamma^*$ for the linear model. **C:** Curvature (absolute value of the second derivative) of the accuracy profiles in panel **B**, evaluated at their peak, as a function of $\mathcal{S}/\sqrt{h}$. The curvature, and hence sensitivity, of the nonlinear model is higher for intermediate and large values of $\mathcal{S}/\sqrt{h}$. Since the functions in panel **B** do not depend on the actual values of $\tilde{h}$ and $\tilde{\gamma}$, but rather the relative distance of these parameters from reference values, what we show in this plot are *relative* curvatures. We compare relative curvatures as $\tilde{h}$ and $\tilde{\gamma}$ do not have the same units. **D:** Ratio of the accuracy of the linear model to that of the normative model, as SNR is varied. Along each curve, the discounting rate $\tilde{\gamma}$ of the linear model is held fixed at the value γ^* that would maximize accuracy at the reference SNR indicated by the legend.

5 | A LINEAR APPROXIMATION OF THE NORMATIVE MODEL

Following Piet et al. (2018) we next show that an approximation of the normative model given by Eq. (2) can be tuned to give near optimal accuracy, but the accuracy of the approximation tends to be sensitive to the changes in the discounting parameter. This approximate, linear model is given by,

$$\frac{dy_t}{dt} = \kappa \left[\sum_{i=1}^{\infty} \delta(t - t_i^R) - \sum_{j=1}^{\infty} \delta(t - t_j^L) \right] - \gamma \cdot y_t. \tag{7}$$

In particular, here the nonlinear sinh term in Eq. (2) is replaced by a linear term proportional to the accumulated evidence.

When tuned appropriately, Eq. (7) closely approximates the dynamics and accuracy of the optimal model (Fig. 3A) (Piet et al., 2018; Veliz-Cuba et al., 2016). Moreover, it also provides a good fit to the responses of rats on a dynamic clicks task (Piet et al., 2018). As the normative and linear models exhibit similar dynamics, it appears that they are difficult to distinguish. However, as we show next, the linear model is more sensitive to changes in its discounting parameter, providing a potential way to distinguish between the models.

We assume that T is large enough so that accuracy has saturated (as in Fig. 1C), and compare the maximal accuracy of the nonlinear and linear model. For the linear discounting rate that maximizes accuracy, $\gamma = \gamma^*$, the linear model

obtains accuracy nearly equal to the normative model (Fig. 3A, inset). The optimal linear discounting rate, γ^*, increases with SNR (Fig. 3A), whereas the discounting term in the normative, nonlinear model remains constant when the hazard rate, h, is fixed. When SNR is large, evidence discounting in the linear model can be stronger (larger γ^*), since each evidence increment is more reliable and can be given more weight. When SNR is lower, linear evidence discounting is weaker (smaller γ^*) resulting in the averaging of noisy evidence across longer timescales.

What is the impact of using the wrong (suboptimal) evidence discounting rate in the two models? To answer this question we compare the accuracy of two observers, one using the nonlinear model with a wrong hazard rate, $\tilde{h} \neq h$, and the other using the linear model with a suboptimal discounting rate $\tilde{\gamma} \neq \gamma^*$. As shown in Fig. 3B accuracy is more sensitive to relative changes in $\tilde{\gamma}$ in the linear model, than relative changes in the assumed hazard rate, $\tilde{h}$, in the nonlinear model. We quantified the sensitivity of both models to changes in evidence discounting rates by computing the curvature of accuracy functions at the optimal discounting value over a range of SNRs (Fig. 3C).

Both models are insensitive to changes in their discounting parameter at low SNR (bottom curve of Fig. 3B). This result is intuitive, as when SNR is small observers perform poorly regardless of their assumptions. On the other hand, when SNR is high observers receive strong evidence from a single click, and the nonlinear model adequately adapts across a broad range of discounting parameter values. The linear model, however, is still sensitive to changes in the discounting parameter, γ. At high SNR, the belief, y_t, as descried by either model is driven to larger values. Whereas the nonlinear model can rapidly discount extreme beliefs as it includes a supralinear leak term, the linear model is not as well adapted, and requires fine tuning. Note, however, that at values of SNR higher than the ones used in Fig. 3B, when, for instance, a single click is sufficient for an accurate decision, both the linear and nonlinear models are insensitive to changes in their discounting parameters. We also note that the insensitivity of the nonlinear model to changes in the discounting rate, h, suggests that this is a more robust model: An observer who does not learn the hazard rate, h, exactly can still perform well. A linear model requires finer parameter tuning to achieve maximal accuracy.

The nonlinear model obtains maximal accuracy as long as the assumed hazard rate matches the true hazard rate $\tilde{h} = h$. On the other hand, the optimal discounting rate of the linear model is also sensitive to changes in the SNR due to changes in the click rates. To quantify this effect, we computed the ratio between the maximal accuracy of the linear model with discounting rate $\tilde{\gamma}$ to the maximal accuracy of the nonlinear model with $\tilde{h} = h$ as the SNR was varied, but h was kept fixed (Fig. 3D). To compute the maximal accuracy we kept $\tilde{\gamma}$ fixed at γ^*, the optimal discounting rate for a reference SNR. The maximal-accuracy ratio for the linear model decreases as SNR changed from this reference SNR, as the optimal discounting parameter of the linear model depends on SNR, and the hazard rate h. Thus, the linear model can achieve maximal accuracy very close to that of the nonlinear model, but this requires fine tuning.

This points to a general difficulty in distinguishing models subjects could use to make inference: Simpler approximations may predict performance that is near identical to that of a normative model. However, this may require precise tuning of the approximations. If the parameters of the task are changed to differ from those on which the subjects have been trained, *i.e.* on tasks where subjects are lead to assume incorrect parameters, the normative model may behave differently from the approximations. In the case we considered, the models may be distinguishable if an animal is extensively trained on trials with fixed parameters h and $\mathcal{S}$, but subsequently interrogated using occasional trials with different task parameters.

The preceding point is illustrated by the following thought experiment. Assume a subject is extensively trained on a fixed set of task parameters: $\mathcal{S}^{\text{ref}} = 2, (\lambda_{\text{low}}, \lambda_{\text{high}}) = (2, 8.5)\text{Hz}, h = 1\text{Hz}$ (peak of the red curve in Fig. 3D). We then introduce some trials with different click rates, say $\mathcal{S}^{\text{new}} = 5$ with $(\lambda_{\text{low}}^{\text{new}}, \lambda_{\text{high}}^{\text{new}}) = (12, 53)\text{Hz}$ and $h = 1\text{Hz}$, chosen so that $\kappa = \log(\lambda_{\text{high}}/\lambda_{\text{low}})$ is constant across the two conditions. We denote by $\text{Acc}_{\text{lin}}(\mathcal{S})$ and $\text{Acc}_{\text{norm}}(\mathcal{S})$ the accuracies of an observer using the linear and normative models on trials with a given $\mathcal{S}$. Since the subject was trained on click rates that correspond to $\mathcal{S}^{\text{ref}}$, their discounting strategy will be adapted to these values. Note that the ratio between $\text{Acc}_{\text{norm}}(\mathcal{S})$

and $\text{Acc}_{\text{lin}}(\mathcal{S})$ when $h = 1$ is the red curve in Fig. 3D. Since the ratio between $\text{Acc}_{\text{norm}}(\mathcal{S}^{\text{ref}})$ and $\text{Acc}_{\text{lin}}(\mathcal{S}^{\text{ref}})$ is near 1, the linear and normative models cannot be distinguished at $\mathcal{S}^{\text{ref}}$. However, a subject using the normative model tuned at $\mathcal{S}^{\text{ref}}$, will still perform optimally at $\mathcal{S} \neq \mathcal{S}^{\text{ref}}$, if κ and h are held constant. On the other hand, a linear model optimized at $\mathcal{S}^{\text{ref}}$, will no longer be optimal at $\mathcal{S} \neq \mathcal{S}^{\text{ref}}$. This distinction is captured by the drop in the accuracy ratio along the red curve in Fig. 3D.

We can quantify the distinction between the two models by their relative difference:

$$\frac{\text{Acc}_{\text{norm}}(\mathcal{S}^{\text{new}}) - \text{Acc}_{\text{lin}}(\mathcal{S}^{\text{new}})}{\text{Acc}_{\text{norm}}(\mathcal{S}^{\text{new}})} = 0.05.$$

More generally, for any decision making model, we may define the quantity

$$D_{\text{model}}(\mathcal{S}) := \frac{\text{Acc}_{\text{norm}}(\mathcal{S}) - \text{Acc}_{\text{model}}(\mathcal{S})}{\text{Acc}_{\text{norm}}(\mathcal{S})},$$

which will equal 0 if the model used is the normative one. If we compute $D_{\text{model}}(\mathcal{S})$ using responses from a real subject, one can generate curves such as those in Fig. 3D. If the curves are not constant (equal to 1), this would suggest the subject is not using an optimal model. Furthermore, a single value of $\mathcal{S}^{\text{new}}$ for which $D_{\text{model}}(\mathcal{S}^{\text{new}}) \neq 0$ provides evidence that the model is not optimal.

In the next two sections, we show how the linear and nonlinear model with added sensory noise differ when fitting the discounting parameters to choice data.

6 | FITTING DISCOUNTING PARAMETERS IN THE PRESENCE OF SENSORY NOISE

The models we have discussed so far translate sensory evidence into decisions deterministically, and do not account for the nervous system's inherent stochasticity (Faisal et al., 2008). We next asked whether the inclusion of sensory noise leads to further differences between the two models, particularly when fit to choice data.

Brunton et al. (2013) showed that in the static version of the clicks task humans and rats make decisions that are best described by a model in which evidence obtained from each click is variable. In the dynamic version of the task, Piet et al. (2018) showed that rats' suboptimal accuracy is well explained by a model that includes similar internal variability. Piet et al. (2018) modeled such "sensory" noise either by applying Gaussian perturbations to the evidence pulses, or by attributing, with some mislocalization probability, a click coming from the right or left speaker to the wrong side.

As a minimal model of neural or sensory noise, we too introduced additive Gaussian noise into the evidence pulse of each click, so that the nonlinear model in Eq. (2) takes the form

$$\frac{dy_t}{dt} = \sum_{i=1}^{\infty} \eta_i \delta\left(t - t_i^R\right) - \sum_{j=1}^{\infty} \zeta_j \delta\left(t - t_j^L\right) - 2h\sinh(y_t), \tag{8}$$

where $\eta_i, \zeta_j \sim \mathcal{N}(\kappa, \sigma)$ are i.i.d. Gaussian random variables with mean κ and standard deviation σ. Similarly, the linear model from Eq. (7) becomes:

$$\frac{dy_t}{dt} = \sum_{i=1}^{\infty} \eta_i \delta\left(t - t_i^R\right) - \sum_{j=1}^{\infty} \zeta_j \delta\left(t - t_j^L\right) - \gamma y. \tag{9}$$

Before fitting these models to choice data, we note that an increase in sensory noise, σ, decreases the value of the discounting parameters that maximize accuracy in both models (Piet et al., 2018): Noisier observations require integration of information over longer timescales (Fig. 4A,B). Thus, adaptivity to change points is sacrificed in order to pool over larger sets of observations . This, in turn, leads to larger biases, particularly after change points. A similar trade-off between adaptivity and bias has been observed in models and human subjects performing a related dynamic decision task (Glaze et al., 2018).

We next fit the discounting parameters in both models using synthetic choice data, treating the other parameters of the models as known. To do so we produced responses using a fixed *reference model* from both classes, and fit a model from each class to the resulting datasets. Specifically, let $m_{\text{ref}} \in \{L, NL\}$ (L = linear, NL = nonlinear) denote the reference model used to produce the choice data, and let $m_{\text{fit}} \in \{L, NL\}$ denote the model that was fit to the resulting data. We independently studied the four possible model pairs ($m_{\text{fit}}, m_{\text{ref}}$). In what follows, θ refers to the discounting parameter that was fit to data in any given class, so that $\theta := \gamma$ when $m_{\text{fit}} = L$ and $\theta := h$ when $m_{\text{fit}} = NL$. Note also that the hazard rate parameter, $\theta = h$, that was fit to data in the case $m_{\text{fit}} = NL$ is distinguished both from the hazard rate, h_{stim}, used to generate click stimuli, and the hazard rate, h_{ref}, used to produce the reference choice data of the nonlinear model. Therefore, to remove ambiguity, we denote by $h_{\text{ref}}, \gamma_{\text{ref}}$ the two constant discounting parameter values used to produce
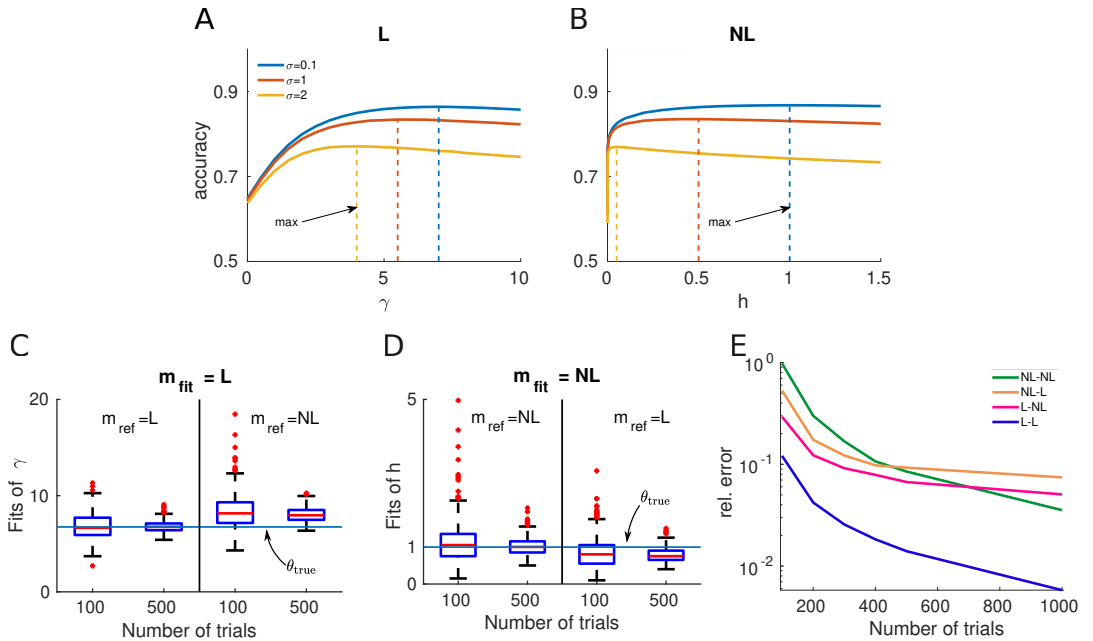

FIGURE 4 **A:** Accuracy as a function of the discounting parameter γ for the linear model with sensory noise described by Eq. (9). As noise increases, maximal accuracy is achieved at lower discounting values (dotted lines). **B:** Same as A for the nonlinear model with discounting parameter h. **C,D** The whisker plots represent the spread of the posterior modes (MAP estimates) obtained across the 500 fitting procedures, for each model pair and reference dataset size. On each box, the red line indicates the median estimate. The MAP estimates are closer with more trials, but are biased in the case of model mismatch. **E:** Average relative error, Eq. (11), in fitting the discounting parameter as a function of reference dataset size. Each color corresponds to a specific pair ($m_{\text{fit}}, m_{\text{ref}}$). In the case of model mismatch (L-NL and NL-L curves) the relative error will not converge to zero due to the bias in the parameter estimate. Panels C-E describe the same set of fits. See Appendix F for simulation details.

the reference choice data with the nonlinear and linear models, respectively. To pick these constants in our simulations, we took the values that would maximize accuracy in the corresponding noise-free systems. That is, $h_{ref} := h_{stim} = 1$ and $\gamma_{ref} := \gamma^* \approx 6.75$ (See Appendix F for more details on the simulations).

During a single fit, we generated stimulus data for N i.i.d. trials,

$$\mathfrak{D} := \{(\mathcal{T}_k, d_k) : 1 \leq k \leq N\}, \tag{10}$$

where $\mathcal{T}_k := \left(\{t_i^R\}_{i=1}^{N_{R,k}}, \{t_j^L\}_{j=1}^{N_{L,k}}\right)$ is the sequence of $N_{R,k}$ right clicks and $N_{L,k}$ left clicks on trial k, and $d_k \in \{0, 1\}$ is the choice datum for this trial. We used Bayesian parameter estimation (See Appendix D.3 for details) to obtain a posterior probability distribution over the discounting parameter, $P(\theta|\mathfrak{D})$.

To account for the variability in the posteriors that arise due to finite size effects, we performed $M = 500$ independent fits per model pair (m_{fit}, m_{ref}), with different dataset sizes: $N \in \{100, 200, 300, 400, 500\}$. To quantify the goodness of these fits, we used the relative mean posterior squared error, averaged across the M fits,

$$\text{err}(p_1, \ldots, p_M, \theta_{true}) := \frac{1}{\theta_{true}^2 M} \sum_{i=1}^{M} \int_0^\infty p_i(\theta)(\theta - \theta_{true})^2 d\theta. \tag{11}$$

This quantity provides a relative measure of how close the posterior distribution is to θ_{true}. Here $p_i(\theta)$ denotes the posterior density, $\Pr(\theta \mid \mathfrak{D})$, from fit i. Note, the definition of θ_{true} is nuanced. If the reference and fit models are the same, then θ_{true} is set to the ground truth, *i.e.* the discounting parameter value used to produce the reference choice data (e.g., $\theta_{true} = h_{ref}$ when $m_{fit} = m_{ref} = $ NL). However, when the fit and reference model classes differ (i.e. when $m_{fit} \neq m_{ref}$), then there is no obvious ground truth, and θ_{true} must be defined differently. In this case, we used the correspondence $\gamma_{ref} \leftrightarrow h_{ref}$. That is, when fitting the nonlinear model we always set $\theta_{true} := h_{ref}$, and when fitting the linear model, we always set $\theta_{true} := \gamma_{ref}$. There are other possible ways of defining θ_{true} in this case, such as picking a discounting parameter value for the fit model class that produces the same accuracy as the reference model. Although arbitrary, our definition is sufficient to illustrate – as we show next – that cross-model fits are feasible and that the $m_{fit} = $ L case is qualitatively different than the $m_{fit} = $ NL case, regardless of the reference model class. However, due to the model mismatch, we expect a bias in the parameter estimate for these situations (i.e. an error that does not converge to 0), unless we define the ground truth self-referentially as the value of the parameter for which the estimate is unbiased.

The maximizer θ^* of our Bayesian posterior $P(\theta|\mathfrak{D})$ defines the maximum likelihood estimate (MLE)[6] of our discounting parameter. We plot the distribution of these across the 500 independent fits, for each (m_{fit}, m_{ref}) pair in Fig. 4C,D. As the number of trials used in the reference dataset increased from 100 to 500, the spread of the estimates diminishes. However, a bias in the estimate appears whenever $m_{fit} \neq m_{ref}$. For reference datasets of size 500, 98% of the 500 MAP estimates in the L-NL fits lie strictly above γ_{true}, versus 50.4% for the corresponding L-L fits. Similarly, 86.6% of the estimates in the NL-L fits lie strictly below h_{true}, versus 44.2% for the corresponding NL-NL fits.

We found that the relative error from Eq. (11) decreases as larger blocks of trials are used to fit the discounting parameter (Fig. 4E). We note the following parallels between the sensitivity to parameter perturbation of each model class (explored in Fig. 3B,C) and the decreasing rate of the relative errors for each model pair. As expected, a model that produces responses that are less sensitive to changes in its discounting parameter requires more trials to be fit to data: The reduction in relative error is the slowest for the (NL, NL) and (NL, L) pairs. This is consistent with the insensitivity of the nonlinear model to changes in discounting parameter, making it difficult to identify its parameters. On the other

---

[6]Which is equal to the Maximum A Posteriori (MAP) estimate in our case, as we picked a uniform prior over a wide interval (See Appendix D.3 for details).

hand, the linear model fits – (L, NL) and (L, L) – converge more rapidly, likely because the linear model is sensitive to changes in its discounting parameter (See Fig. 3B,C).

In anticipation of our next section, we point out that computing the MLE can be treated as a statistical learning problem in which we minimize a negative log-likelihood loss function over the dataset $\mathfrak{D}$ (See Eq. 7.8 in Friedman et al. (2001)):

$$\mathcal{L}_{LL}(d_k, m_{\text{fit}}(\mathcal{T}_k)|\sigma, \theta) := -\log \mathsf{P}(d_k = m_{\text{fit}}(\mathcal{T}_k)|\sigma, \theta). \tag{12}$$

Here d_k and $m_{\text{fit}}(\mathcal{T}_k)$ are the choices generated by the reference and fit models, respectively, on the kth trial. As before the discounting parameter, θ, and the level of sensory noise, σ, parametrize the fitted model. Fitted model responses $m_{\text{fit}}(\mathcal{T}_k)$ are non-deterministic only because of sensory noise. The likelihood $\mathsf{P}(d_k = m_{\text{fit}}(\mathcal{T}_k)|\sigma, \theta)$ is the probability that the response generated by the fit model on trial k matches the response observed in the data (See Appendix F for details on how this likelihood was computed for each model class), which must be obtained from many realizations of $m_{\text{fit}}(\mathcal{T}_k)$ subject to click noise of amplitude σ. The MLE, θ^*, for m_{fit} is then found by minimizing the expected loss across all trials,

$$\theta^* := \operatorname{argmin}_\theta \mathsf{E}\left[\mathcal{L}_{LL}(d_k, m_{\text{fit}}(\mathcal{T}_k)|\sigma, \theta)\right] = \operatorname{argmin}_\theta \frac{1}{N} \sum_{k=1}^{N} \mathcal{L}_{LL}(d_k, m_{\text{fit}}(\mathcal{T}_k)|\sigma, \theta), \tag{13}$$

taking the expectation over all N samples in the dataset, but conditioning on the fitted model's discounting parameter, θ and noise amplitude, σ. As the MLE is consistent, we expect the fit parameters will converge to the true parameters (Wald, 1949) (Fig. 4C,D). Framing Bayesian parameter estimation in this way will help us compare to our approach of fitting by minimizing the 0/1-loss function we introduce next.

7 | FITTING WITH THE 0/1-LOSS FUNCTION

We next asked how the parameters that define the model whose responses best match the choices of a reference observer compare to those that maximize the likelihood of observing these choices. As we noted minimizing the log-likelihood loss $\mathcal{L}_{LL}$ given in Eqs. (12) and (13) gives the parameters most likely to have produced the data, and we expect the corresponding estimates of the discounting parameter to converge to the true value when the fit and reference models match.

To find the parameters that maximize the probability of matching the choices of the model to those observed in a dataset on every trial, we define the 0/1-loss function,

$$\mathcal{L}_{0/1}(d_k, m_{\text{fit}}(\mathcal{T}_k, Z_j)|\sigma, \theta) := \mathbb{1}(d_k \neq m_{\text{fit}}(\mathcal{T}_k, Z_j)|\sigma, \theta),$$

where $\mathbb{1}$ is the indicator function, $(\mathcal{T}_k, d_k)$ is a data sample indicating the click stimulus and response on a trial k, and $m_{\text{fit}}(\mathcal{T}_k, Z_j)$ is the response of the fitted model with discounting parameter θ, click stimulus $\mathcal{T}_k$, and $Z_j, j = 1, ..., Q$ are realizations of sensory noise, *i.e.* a sequence of i.i.d. Gaussian variables that perturb the evidence obtained from each click. We will marginalize over realizations of the (unobserved) sensory noise, and Q denotes the number of realizations we use in the actual computation. Fitting the discounting parameter θ then involves minimizing the empirical expectation

of the loss function $\mathcal{L}_{0/1}$ over the data samples $(\mathcal{T}_k, d_k)$ and across realizations, Z_j, of sensory noise,

$$\theta^* := \text{argmin}_\theta \text{E}\left[\mathcal{L}_{0/1}(d_k, m_{\text{fit}}(\mathcal{T}_k, Z_j)|\sigma, \theta)\right] = \text{argmin}_\theta \frac{1}{QN} \sum_{j=1}^{Q} \sum_{k=1}^{N} \mathcal{L}_{0/1}(d_k, m_{\text{fit}}(\mathcal{T}_k, Z_j)|\sigma, \theta).$$

For a binary decision model, this involves finding the parameter θ that minimizes the expected number of mismatches (or probability of a mismatch) between the choices of the model and those observed by the data (minimizing 0/1-loss), or maximizes the expected number of matches (or probability of a match) between the data and fit model (maximizing 0/1-prediction accuracy). In our fits, we used $Q = 1$, sampling a single realization of click noise perturbations per click stream. As we sampled from a large number of click streams, this was sufficient to average the loss function.

Both loss functions, $\mathcal{L}_{LL}$ and $\mathcal{L}_{0/1}$, are commonly used to fit models to data (Friedman, 1997; Friedman et al., 2001). Minimizing the expectation of $\mathcal{L}_{0/1}$ is reasonable, as it seems likely that the parameters that define the model that matches the largest number of choices observed in the data should be close to the one the reference observer actually uses (assuming that there is no model mismatch). These parameters will then also best predict future responses. On the other hand, minimizing $\mathcal{L}_{LL}$ produces the most likely parameters that produced the observed data.

However, it is well-known that parameters estimated using different cost functions can differ, even when the models used to fit and generate the data agree. To see the difference between using $\mathcal{L}_{LL}$ and $\mathcal{L}_{0/1}$ in Eq. (13) consider a Bernoulli random variable, B with parameter $p > 0.5$. Given a large sequence of observed outcomes, $N \to \infty$, the parameter that minimizes the expected loss $\bar{\mathcal{L}}_{LL}$ converges to p, as the MLE is consistent and asymptotically efficient (Wald, 1949). On the other hand, the parameter that minimizes the expected loss $\bar{\mathcal{L}}_{0/1}$ is $p = 1$ (See Appendix E): The individual outcomes in a series of independent trials are best matched by a model that always guesses the more likely outcome.

We observed a similar bias when we used the $\mathcal{L}_{0/1}$ loss function to infer the discounting parameters in our evidence accumulation models (Friedman, 1997): We generated a set of 10^6 click-train realizations, $\mathcal{T}_k$, and two sets of responses, d_k, from each the linear and nonlinear evidence accumulation models with sensory noise (See Appendix F). Next we used these stimulus realizations as input to an evidence-accumulation model (linear or nonlinear) with a fixed discounting parameter to produce a corresponding set of 10^6 *reference observer responses*. We generated a second set of *model responses* using the same database of 10^6 click-train realizations, but allowed the discounting parameter θ to vary. We call the fraction of time the reference observer and model responses agree the *0/1-prediction accuracy* (PA) of the model, the complement of the expected 0/1-loss over a test set, PA $= 1 - \bar{\mathcal{L}}_{0/1}$. When the model and reference observer agree the PA is 1 in the absence of sensory noise ($\sigma = 0$), as the stimulus determines the choice fully. However, PA decreases as sensory noise increases.

Somewhat surprisingly, the parameters that minimize expected 0/1-loss are biased, and this bias increases with sensory noise (Fig. 5). In particular, the discounting parameter that best predicts the reference observer responses is lower than the one used to generate these responses (Fig. 5B,D). This is consistent with our observations in Section 6, as integration over longer periods of time decreases response variability (Fig. 4A,B). This tendency is pronounced when larger values of the discounting parameters are used to produce the training data. Larger discounting leads to shorter integration time, and increased variability in the responses. Furthermore, the nonlinear (NL) model exhibits this bias much more strongly than the linear model (L). See Appendix G for a possible metric of the reported bias, and its dependence on sensory noise for each model class (Fig. 6).

Thus sensory noise is the main reason the expected 0/1-loss is minimized at a discounting parameter that does not match the one used to generate the data. Such internal noise introduces variability in the responses: even the same model will not match its own responses given the same stimulus, and a decrease in output variability can increase the PA of a model. In the present case, such a decrease in response variability is achieved by decreasing the discounting
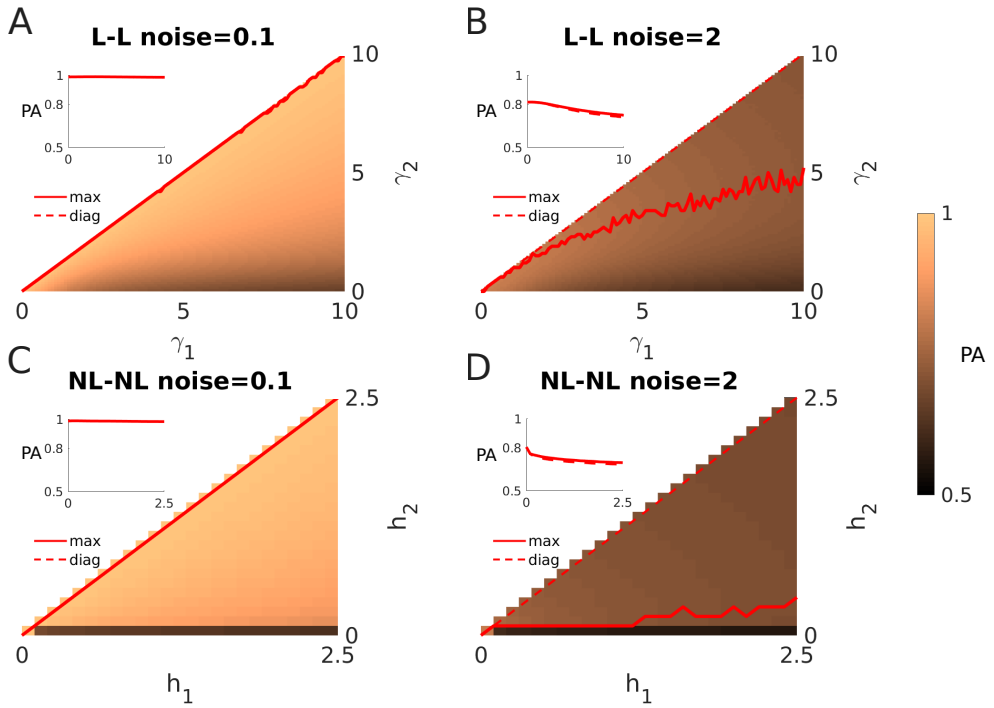
FIGURE 5 The 0/1-prediction accuracy (PA) of a fit model as a function of the discounting parameter of the fit model (vertical axis), and model used to generate training data (horizontal axis). Each column corresponds to a noise value, and each row to a model (linear or nonlinear). Colors indicate the variation of the PA as the discounting parameters are varied (for $\gamma_2 \leq \gamma_1$, as PA is symmetric about the diagonal). Red curves represent the fit parameter that maximizes percent match, as a function of the reference parameter. For higher noise values, this lies well below the diagonal, $\gamma_1 = \gamma_2$, which would correspond to matching the parameters of the reference model and data. The same 10^6 click realizations were used across all panels, but each decision from the two models was computed with independent sensory noise realizations. Other parameters are $h_{\text{stim}} = 1$ Hz, $(\lambda_{\text{high}}, \lambda_{\text{low}}) = (20, 5)$ Hz, and $T = 2$ s.

parameter, and increasing integration time.

We expect that similar biases occur whenever a 0/1-loss function is used to fit models to choice data. Sensory noise, lapses in attention, and numerous other sources of noise nearly always introduce some variability in the responses of observers. In such cases, models that are less variable than the observer may best match an observed set of responses, and best predict future responses (Friedman et al., 2001). However, these parameters are not always most likely to have been used by the observer. Using a 0/1-loss function may thus not always reveal the process that the observer used to generate the responses, even if the model the observer uses is close to the one used to fit the data.

8 | DISCUSSION

Normative models of decision-making make concrete predictions about the computations and actions of experimental subjects, and can be used to interpret behavioral data (Geisler, 2003). Such models can also be used to identify task parameter ranges in which observers' responses are most sensitive to their assumptions about the task. In turn, such

information can then be used to tease apart candidate model classes the experimental subject might be employing. Here we have focused on properties of a normative, nonlinear model, and its differences with a close, linear approximation. We found that the linear model is more sensitive to changes in the discounting parameter compared to the nonlinear model, and suggest this is why fitting a linear model to choice data requires fewer trials than fitting a nonlinear model.

In dynamic environments, task parameters may have predictable effects on subjects' overall accuracy and accuracy relative to change points. We have shown that there is a range of intermediate to high SNR in which the linear model is sensitive to changes in its discounting parameter, but the nonlinear model is not. This suggests this range could be probed to distinguish the evidence accumulation strategies subjects are using. These strategies may also be fit by other approximate models, like accumulators with no-flux boundaries or sliding-window integrators (Wilson et al., 2013; Glaze et al., 2015; Barendregt et al., 2019), which can also be sensitive to changes in their discounting parameters.

Psychophysical tasks used to infer subjects' decision-making strategies can require extensive training and data collection (Hawkins et al., 2015b,a). Normative and approximately normative decision-making models diverge most in their response accuracy when tasks are of intermediate difficulty. As we have shown, task difficulty may be controlled by combinations of task parameters representing fewer dimensions than the total number of parameters. Identifying these parameter combinations may be possible by computing the signal-to-noise (SNR) ratio of the stimulus produced by a particular parameter set. However, subjects' responses are also susceptible to noise from sensing and processing evidence, so it is important to extend descriptions of SNR to account for such factors (Brunton et al., 2013).

Normative models of evidence accumulation and decision-making can be complex, and simpler, approximately optimal strategies may perform nearly as well (Wilson et al., 2010; Glaze et al., 2018). If such approximate strategies are easier to learn and tune, subjects may prefer them. Piet et al. (2018) showed rats' performance on the dynamic clicks task is well fit by a linear discounting model. However, optimal and well-tuned suboptimal strategies may be difficult to distinguish, and this problem is likely to worsen with increasing task complexity and corresponding model dimensionality. We have described possible model-guided task design approaches that may help tease apart similarly performing models.

The addition of noise in our evidence accumulation models provides an extra parameter that can account for suboptimal performance. What is the best way to distinguishing whether internal noise or suboptimal evidence accumulation strategies best account for underperformance? One way to do this, as suggested by our model analysis, is to collect sufficient data over trials in which a task parameter was changed unbeknownst to the observer.

For purposes of model fitting to experimental data, we expect that trial-to-trial variability can be more faithfully tracked in the dynamic clicks task than in dynamic decision tasks based on the RDMD task. This is due to the relative simplicity of the clicks as evidence sources: They are either on the right or left, although click side and evidence strength can be misattributed (Piet et al., 2018). In contrast, dot motion can be estimated in many ways, making it difficult to interpret which aspects of the stimulus an animal observed, and used as evidence. Spatiotemporal sampling methods may be too spatially coarse and may require fitting filters to each subject, which could change trial-to-trial (Adelson and Bergen, 1985; Park et al., 2016). Transforming click times to delta pulses using Eq. (2) is more straightforward. Thus, the dynamic click task paradigm is a promising avenue for probing evidence accumulation to complement dynamic tasks which are extensions of classic RDMD (Glaze et al., 2015).

The use of discrete evidence tasks does come with caveats. The neural computations underlying visual motion discrimination in non-human primates are well studied (Born and Bradley, 2005), and have a significant history of being linked to decision tasks (Gold and Shadlen, 2007). As a result, there is an extensive literature connecting neural systems for processing visual motion and those involved in decision deliberation (Shadlen and Newsome, 1996; Roitman and Shadlen, 2002). However, only recently have the neural underpinnings of the decisions of rats performing auditory discrimination tasks been examined (Brody and Hanks, 2016). Furthermore, mathematical issues may arise in precisely

characterizing discounting between clicks, when evidence arrives discretely. Many different functions could lead to the same amount of evidence discounting between clicks, leading to ambiguity in the model selection process.

Parameter identification for evidence accumulation models can be sensitive to the method chosen to fit model responses to choice data (Ratcliff and Tuerlinckx, 2002). Glaze et al. (2015) used the approach of minimizing the cross-entropy error function, which measures the dissimilarity between binary choices in the model and the data. Piet et al. (2018) used a maximum likelihood approach to identify model parameters that most closely matched choice data. This is related to the Bayesian estimation approach we used to fit parameters of the nonlinear and linear models. We obtained similar results by minimizing the expected 0/1-loss, which biases towards less variable models, especially for models with strong sensory noise (Fig. 5). A more careful approach to fitting model parameters should also consider penalizing more complex models, which would also allow us to distinguish between the nonlinear and linear model.

Glaze et al. (2018) recently studied the strategies humans use when making binary decisions in dynamic environments whose hazard rates changed across trial blocks. In this case the ideal observer must infer both the state, and the rate at which the environment is changing (Radillo et al., 2017). Interestingly, Glaze et al. (2018) found that the model that best accounted for response data was not the full Bayes optimal model, but rather a sampling model in which a bank of possible hazard rates replaces the full hazard rate distribution. Such sampling strategies can more easily be implemented in spiking networks (Buesing et al., 2011), and may also arise when considering an information bottleneck, which forces a balance between information required from the past and model predictivity (Bialek et al., 2001). As in Occam's razor, the brain may favor simpler models, especially when they perform similarly to more complex models (Balasubramanian, 1997).

Analyses of normative models for decision-making are important both for designing experiments that reveal subjects' decision strategies and for developing heuristic models that may perform near-optimally (Veliz-Cuba et al., 2016; Piet et al., 2018; Glaze et al., 2018). Our findings suggest subjects should be tested mainly at intermediate levels of SNR to provide informative response data. We found that such a level of SNR is between 1 and 2 for an optimal observer, and between 3 and 4 for an observer that uses linear discounting. Tasks that are too easy or hard allow subjects to obtain similar performance with a wide variety of strategies. Interestingly, we also found that the models that best predict observer responses, are not necessarily those closest to the ones that the observer is using. Moreover, modifications of normative models can also suggest more revealing experiments, like those that include feedback or signaled change points. Ultimately, data from decision-making tasks that require subjects to accumulate evidence adaptively will provide a better picture of how organisms integrate stimuli to make choices in the natural world.

A | NORMATIVE EVIDENCE ACCUMULATION FOR DYNAMIC CLICKS

In dynamic environments, the state $x_t \in \{x^R, x^L\}$ evolves according to a continuous-time Markov chain with symmetric transition rates given by the hazard rate, h. We construct a sampled-time approximation $\{\bar{x}_t^{\Delta t}\}_{t \in [0,T]}$ of the continuous-time Markov process x_t, parameterized by Δt, which is valid for Δt small enough (Gardiner, 2009). More precisely, we define a discrete-time Markov chain $x_n \in \{x^R, x^L\}$ by the transition probabilities: $P(x_n \neq x_{n-1}|x_{n-1}) = h \cdot \Delta t$ and $P(x_n = x_{n-1}|x_{n-1}) = 1 - h \cdot \Delta t$, for all $n \in \mathbb{N}$ and initial condition x_0. Note that these probabilities are a truncation to first order in Δt of the transition probabilities that one would otherwise obtain for the embedded discrete-time Markov chain $\{x_{n\Delta t}\}_{n \in \mathbb{N}}$. Then, we set

$$\bar{x}_t^{\Delta t} := x_n, \tag{14}$$

for all $n \in \mathbb{N}$ and all $n\Delta t \leq t < (n+1)\Delta t$. In the following, our discrete-time evidence accumulation equations are embedded in continuous-time via the correspondence given by Eq. (14). As $\Delta t \to 0$ the resulting equations apply to the original state process x_t in virtue of the sampled-time approximation just described.

Just as in Eq. (1), the log-posterior odds ratio in discrete-time is:

$$y_n := \log \frac{P(x_n = x^R | s^R(n), s^L(n))}{P(x_n = x^L | s^R(n), s^L(n))}.$$

Hence, equations (A.3) and (B.1) from the appendix of Veliz-Cuba et al. (2016) hold in our context:

$$y_n - y_{n-1} = \log \frac{f_{\Delta t}^R(\xi_n)}{f_{\Delta t}^L(\xi_n)} + \log \frac{1 - h \cdot \Delta t + h \cdot \Delta t \cdot e^{-y_{n-1}}}{1 - h \cdot \Delta t + h \cdot \Delta t \cdot e^{y_{n-1}}}.$$

In addition, we use the approximation $\log(1+z) \approx z$ for small $|z|$, since $0 < \Delta t \ll 1$, so that:

$$\Delta y_n = \log \frac{f_{\Delta t}^+(\xi_n)}{f_{\Delta t}^-(\xi_n)} - 2h\Delta t \sinh(y_{n-1}).$$

Taking the limit $\Delta t \to 0$ yields the ODE:

$$\frac{dy_t}{dt} = \kappa \left[\sum_{i=1}^{\infty} \delta(t - t_i^R) - \sum_{j=1}^{\infty} \delta(t - t_j^L) \right] - 2h \sinh(y_t),$$

or the equivalent rescaled version

$$\frac{dy_t}{dt} = \sum_{i=1}^{\infty} \delta(t - t_i^R) - \sum_{j=1}^{\infty} \delta(t - t_j^L) - \frac{2h}{\kappa} \sinh(\kappa \cdot y_t),$$

which both appear in Piet et al. (2018).

B | DERIVATION OF DYNAMIC CLICKS SNR

Our derivation considers the *signal* in the dynamic clicks task to be the difference in the number of clicks during the final epoch prior to interrogation at time $t = T$. The distribution of final epoch times τ of the telegraph process $x(t)$ is

$$p(\tau) = he^{-h\tau} + e^{-h\tau}\delta(\tau - T), \qquad \tau \in [0, T]. \tag{15}$$

The first term is the distribution of waiting times between switches. We truncate the period at the interrogation time, T, and the second term accounts for the probability that no switches occur during the entire trial, and the final and only epoch is of length T. For a final epoch of a given length τ, we can describe both the conditional expectation and variance of the difference in click counts ΔN again using the results of Skellam (1946):

$$E[\Delta N | \tau] = (\lambda_{\text{high}} - \lambda_{\text{low}})\tau, \qquad \text{Var}[\Delta N | \tau] = (\lambda_{\text{high}} + \lambda_{\text{low}})\tau.$$

Therefore to obtain the unconditional expectation and variance for ΔN, we must marginalize using the laws of total expectation and variance with respect to the distribution of epoch times τ given in Eq. (15). This yields

$$E[\Delta N] = (\lambda_{\text{high}} - \lambda_{\text{low}}) \int_0^T \tau p(\tau) d\tau = \frac{1 - e^{-hT}}{h} \cdot (\lambda_{\text{high}} - \lambda_{\text{low}}) \tag{16}$$

for the total expectation. Notice that as $T \to \infty$, the expected number of clicks is limited from above by $\lim_{T \to \infty} E[\Delta N] = (\lambda_{\text{high}} - \lambda_{\text{low}})/h$. Using the law of total variance we can thus compute

$$\text{Var}[\Delta N] = \text{Var}\left[E[\Delta N|\tau]\right] + E\left[\text{Var}[\Delta N|\tau]\right] = (\lambda_{\text{high}} - \lambda_{\text{low}})^2 \cdot \text{Var}[\tau] + \frac{1 - e^{-hT}}{h} \cdot (\lambda_{\text{high}} + \lambda_{\text{low}})$$

$$= \frac{1 - 2hT e^{-hT} - e^{-2hT}}{h^2} (\lambda_{\text{high}} - \lambda_{\text{low}})^2 + \frac{1 - e^{-hT}}{h} (\lambda_{\text{high}} + \lambda_{\text{low}}). \tag{17}$$

Plugging Eq. (16) and (17) into the expression for $\text{SNR}_h^T = E[\Delta N]/\sqrt{\text{Var}[\Delta N]}$ yields

$$\text{SNR}_h^T = \frac{(1 - e^{-hT})(\lambda_{\text{high}} - \lambda_{\text{low}})}{\sqrt{(1 - 2hT e^{-hT} - e^{-2hT})(\lambda_{\text{high}} - \lambda_{\text{low}})^2 + h \cdot (1 - e^{-hT})(\lambda_{\text{high}} + \lambda_{\text{low}})}}. \tag{18}$$

Recalling our definition from equation (4),

$$\mathcal{S} := \frac{\lambda_{\text{high}} - \lambda_{\text{low}}}{\sqrt{\lambda_{\text{high}} + \lambda_{\text{high}}}}, \tag{19}$$

we can rewrite equation (18) in the more convenient form

$$\text{SNR}_h^T = F(hT, \mathcal{S}/\sqrt{h}) = \frac{(1 - e^{-hT})\mathcal{S}/\sqrt{h}}{\sqrt{(1 - 2hT e^{-hT} - e^{-2hT})\mathcal{S}^2/h + 1 - e^{-hT}}}, \tag{20}$$

where we have highlighted the fact that the SNR is a function of the rescaled trial time hT and the Skellam SNR rate $\mathcal{S}$ scaled by the root of the timescale $\sqrt{1/h}$. Indeed, in the limit as $h \to 0$, we find that $\text{SNR}_h^T \to \text{SNR}_0^T = \mathcal{S} \cdot \sqrt{T}$ consistent with Eq. (3). We also find that in the limit of infinitely long trials $T \to \infty$, Eq. (20) tends to

$$\lim_{T \to \infty} \text{SNR}_h^T = \frac{\mathcal{S}/\sqrt{h}}{\sqrt{\mathcal{S}^2/h + 1}},$$

so the SNR is solely determined by $\mathcal{S}/\sqrt{h}$.

Note also that to keep Eq. (20) constant it is sufficient to keep its constituent arguments constant. This is convenient, since we already must keep hT constant to fix the statistics of information accumulated prior to the final epoch, so we predict that performance is fixed by the following two parameters

$$\frac{\mathcal{S}}{\sqrt{h}} = \frac{\lambda_{\text{high}} - \lambda_{\text{low}}}{\sqrt{h(\lambda_{\text{high}} + \lambda_{\text{low}})}} = c_1 \quad \text{and} \quad hT = c_2,$$

as reported in Eq. (6).

C | DIFFUSION APPROXIMATION

Here we demonstrate the diffusion approximation of the normative model for the dynamic clicks task, Eq. (2) in the limit of large Poisson rates λ_{high} and λ_{low}. Diffusion approximations for jump processes have been addressed by Lánskỳ (1997), and Richardson and Swarbrick (2010) who studied the impact of shot noise and pulsatile synaptic inputs on integrate-and-fire models. Following this work, we note that the difference of the click streams in Eq. (2) can be approximated by a drift-diffusion process with matched mean, $\pm\kappa(\lambda_{\text{hi}} - \lambda_{\text{low}})$, and variance, $\kappa^2(\lambda_{\text{hi}} + \lambda_{\text{low}})$. This results in the following stochastic differential equation (SDE) for the approximation $\bar{y}_t$:

$$d\bar{y}_t = \kappa g(t)(\lambda_{\text{high}} - \lambda_{\text{low}})dt + \kappa\sqrt{\lambda_{\text{high}} + \lambda_{\text{low}}}dW_t - 2h\sinh(\bar{y}_t)dt, \tag{21}$$

where $g(t) = \text{sign}\left[\lambda_R(t) - \lambda_L(t)\right]$ and dW_t is the increment of a Wiener process. Note the resulting nonlinear drift-diffusion model is similar to the normative models presented in (Glaze et al., 2015; Veliz-Cuba et al., 2016). The SNR of the signal in Eq. (21) can be associated with the mean divided by the standard deviation in an average-length epoch. Fixing this SNR leads to the relations in Eq. (6). Importantly, the signal in Eq. (21) is characterized entirely by its mean and variance, so we expect that the performance of the model can be directly associated with the SNR. Note, however, that Eq. (21) will only be valid for $\lambda_{\text{high}}, \lambda_{\text{low}} \gg 1$. Otherwise, one must consider the effects of higher order moments of the click streams, and a prediction of performance purely based on the SNR will break down (Fig. 1D, Inset), since higher order statistics likely shape response accuracy in these cases.

D | MODEL IDENTIFICATION

We fit parameters of the linear and nonlinear models in two stages. First, we generated synthetic response data from a model (linear or nonlinear) by solving the corresponding ODE or SDE. We then solved a second set of models (linear or nonlinear) for a range of discounting parameters (γ for the linear model; h for the nonlinear model), and constructed a posterior distribution over the discounting parameter. For noisy models, we expect the posterior to be a smooth function that is peaked around the most likely values of discounting parameter for that trial. We now describe the details of these parameter fitting procedures for each of the cases: linear vs. nonlinear models.

D.1 | Linear model with stochastic response

We incorporate noise into the linear Eq. (7) by considering multiplicative noise on the click increments, as described by Eq. (9). For a fixed realization of the click train, we can solve this equation explicitly for y_t at the end of trial k:

$$y_T^k = \sum_{i=1}^{N_R^k} \eta_i e^{-\gamma(T^k - t_R^i)} - \sum_{j=1}^{N_L^k} \zeta_j e^{-\gamma(T^k - t_L^j)}, \tag{22}$$

where $\eta_i, \zeta_j \sim \mathcal{N}(\kappa, \sigma)$, revealing y_{T^k} is simply the sum of i.i.d. normal random variables scaled by exponential decay. Conditioning on the clicks $\mathcal{T}_k$, then y_{T^k} is normally distributed $[y_{T^k} | \mathcal{T}_k, \gamma]$ with expectation and variance

$$\mathcal{E}_k := \mathsf{E}\left[y_{T^k} | \mathcal{T}_k, \gamma\right] = \kappa \left[\sum_{i=1}^{N_R^k} e^{-\gamma(T^k - t_R^i)} - \sum_{j=1}^{N_L^k} e^{-\gamma(T^k - t_L^j)}\right],$$

$$\mathcal{V}_k := \mathrm{Var}\left[y_{T^k} | \mathcal{T}_k, \gamma\right] = \sigma_c^2 \left[\sum_{i=1}^{N_R^k} e^{-2\gamma(T^k - t_R^i)} + \sum_{j=1}^{N_L^k} e^{-2\gamma(T^k - t_L^j)}\right],$$

so $\tilde{r}_k(\gamma) \in \pm 1$ is a Bernoulli random variable. The likelihood function will be a smooth function of γ, and determined as an integral over the half-line corresponding to the ± 1 decision:

$$\mathsf{P}(\tilde{r}_k = \pm 1 | \gamma, \mathcal{T}_k) = \pm \int_0^{\pm\infty} p(y_{T^k} | \mathcal{T}_k, \gamma) \mathrm{d}y_{T^k} = \Phi\left(\pm \frac{\mathcal{E}_k}{\sqrt{\mathcal{V}_k}}\right),$$

where $\Phi(z)$ is the cumulative distribution function of a standard normal random variable. We can thus compute the posterior over the discounting parameter γ as a rescaled product of the likelihoods on each trial.

D.2 | Nonlinear model with stochastic response

When click heights are noise-perturbed, we cannot explicitly solve the extended nonlinear model. However, we can make progress by applying the idea of mapping between clicks. If we draw trains of clicks, $\mathcal{T}_k$, ahead of time, Eq. (8) defines the nonlinear model with multiplicative noise. We can iteratively define the probability density $p(y, t)$ by sampling over the click amplitude noise distribution at each click according to

$$p(y, t_n^+) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(y-z-\kappa_n)^2/2\sigma^2} p(z, t_n^-) \mathrm{d}z, \tag{23a}$$

$$\frac{\partial p(y, t)}{\partial t} = 2h \cdot \frac{\partial}{\partial y}\left[\sinh(y) p(y, t)\right], \qquad t_n < t < t_{n+1}, \tag{23b}$$

where click noise is drawn from the normal distribution $\mathcal{N}(\kappa, \sigma)$, t_n is the time of the n-th click, $\kappa_n = \pm\kappa$ according to the side of the n-th click, and we have used the convolution theorem for independent random variables. For any trains of clicks, $\mathcal{T}_k$, Eq. (23) can be solved iteratively to obtain the distribution $p(y, T)$. The likelihood will thus be a smooth function of h, determined by the integral over the half-line corresponding to the decision (± 1):

$$\mathsf{P}(\tilde{r}_k = \pm 1 | h, \mathcal{T}_k) = \pm \int_0^{\pm\infty} p(y_{T^k} | \mathcal{T}_k, h) \mathrm{d}y_{T^k}. \tag{24}$$

D.3 | Bayesian fitting procedure

Our goal is to compute or estimate the posterior distribution $\mathrm{Pr}(\theta | \mathfrak{D})$, which by Bayes' rule is proportional to the product of the likelihood of the data with the prior over the parameter[7]:

$$\mathsf{P}(\theta | \mathfrak{D}) \propto \mathsf{P}(\mathfrak{D} | \theta) \mathsf{P}_0(\theta). \tag{25}$$

---

[7] Since all the other task and model parameters are assumed known and fixed, we may omit them from the equations.

Our method focuses on exploiting the likelihood function $P(\mathfrak{D}|\theta)$. We have,

$$P(\mathfrak{D}|\theta) = P(\mathcal{T}_{1:N}, d_{1:N}|\theta) = P(d_{1:N}|\mathcal{T}_{1:N}, \theta)P(\mathcal{T}_{1:N}|\theta) = P(d_{1:N}|\mathcal{T}_{1:N}, \theta)P(\mathcal{T}_{1:N}),$$

where the last step comes from the fact that the clicks trains are independent of the discounting parameter θ used by the decision-making model[8]. From there, we remark that the choice data are conditionally independent on the clicks stimulus and the discounting parameter. Thus,

$$P(d_{1:N}|\mathcal{T}_{1:N}, \theta) = \prod_{k=1}^{N} P(d_k|\mathcal{T}_k, \theta).$$

Therefore we can rewrite Eq. (25) as:

$$P(\theta|\mathfrak{D}) \propto P_0(\theta) \prod_{k=1}^{N} P(d_k|\mathcal{T}_k, \theta). \tag{26}$$

We use uniform priors for θ, over a finite interval $[0, a]$. In this context, the problem of computing the posterior distribution of θ reduces to assessing the likelihoods of the decision data on each trial, $P(d_k|\mathcal{T}_k, \theta)$ $(1 \leq k \leq N)$, for a range of θ-values spanning the interval $[0, a]$. In practice, we picked $a = 40$ when fitting the linear model and $a = 10$ when fitting the nonlinear model. Finally, note that for numerical stability reasons, our algorithms actually sum log-likelihood values, as opposed to multiplying probability values. Relegating the θ-independent prior into a normalization constant C, Eq. (26) becomes, in the log-domain:

$$\log P(\theta|\mathfrak{D}) = C + \sum_{k=1}^{N} \log P(d_k|\mathcal{T}_k, \theta), \quad \theta \in [0, a]. \tag{27}$$

E | MINIMIZING 0/1-LOSS IN A BERNOULLI RANDOM VARIABLE

Consider a simple stochastic binary decision-making model in which we ignore the specifics of evidence sources, as in Pesaran and Timmermann (1992). We that in this case the 0/1-loss function also leads to biased estimates. This result has been pointed out in previous work in which parameter fitting results have been compared between Bernoulli random variables fit with the 0/1-loss function as opposed to maximum likelihood estimators (Friedman, 1997; Friedman et al., 2001).

Consider a Bernoulli random variable B_1 with success probability p_1 generating the reference choices, and the fit Bernoulli model B_2 with success probability p_2. Minimizing the log-likelihood loss function recovers $p_2^* = p_1$ in the limit of a large number of trials $N \to \infty$: In this limit, given p_2, we have that the expected loss measured by the negative log-likelihood is

$$\bar{\mathcal{L}}_{LL}(d|p_2) = -[p_1 \log p_2 + (1 - p_1) \log(1 - p_2)], \tag{28}$$

which is minimized[9] at $p_2^* = p_1$, the mean of B_1. Thus, the parameter from the reference model is recovered, as the

---

[8] We remind the reader that we operate a distinction between the discounting parameter of the decision maker and the hazard rate used to produce the data.

[9] Note Eq. (28) is the cross-entropy between B_1 and B_2.

Bernoulli random variable satisfies the requirements for the MLE to be consistent (Wald, 1949).

On the other hand, if we fit the parameter p_2 by minimizing the expected 0/1-loss function, in the limit of $N \to \infty$ trials, the expected loss is

$$\bar{\mathcal{L}}_{0/1}(d|p_2) = 1 - P(B_1 = B_2) = p_1 - p_2(2p_1 - 1),$$

which decreases in p_2 for $p_1 > 0.5$, so the minimal expected loss when $p_1 > 0.5$ is achieved with $p_2 = 1$ (for $p_1 < 0.5$ it is minimized at $p_2 = 0$).

Of course, the synthetic data and the fit evidence accumulation models we consider are generated from the same click streams on each trial, so a realistic comparison should account for such noise correlations in simplified Bernoulli random variable models, as analyzed in Dai et al. (2013). This analysis is more involved, and we save such a study for future work.

F | DETAILS ON MONTE CARLO SIMULATIONS FOR FIGURES

Fig. 1C was generated using 10^5 simulations of Eq. 2 from $t = 0$ to $t = 1$s with the parameters shown in the figure. The time for saturation was chosen to be 0.4s. For each time between 0 and 0.4s the accuracy was computed as the percentage of the 10^5 simulations for which the choices were correct. Fig 1D was generated using 10^5 simulations of Eq. 2 for each data point in the $(\lambda_{\text{high}}, \lambda_{\text{low}})$ plane. The maximal accuracy reported corresponds to the numerically computed accuracy at $t = 1$s.

Fig. 2B-D was generated using 10^5 simulations of Eq. 2 from $t = 0$ to $t = 3s$ with the parameters shown in the figure. The reference change point was chosen to be the last change point in the simulation. For each time between the last change point and one unit of time later, the accuracy is the fraction of the correct responses, simulations for which $\text{sign}(y_T) = x(T)$, the sign of the LLR matched the sign of the telegraph process. Since intervals between change points are exponentially distributed, there are many more data points for short times than for long times after change points. Since some simulations did not last a full unit of time after the last change point, the number of simulations is less than or equal to 10^5 (decreasing as time increases). Simulations that had no change point were omitted when computing the accuracy.

Fig. 3A was generated as follows. For each value of $\mathcal{S}/\sqrt{h}$, 10^6 simulations of Eq. 7 from $t = 0$ to $t = 1s$ were generated over a range of γ values. For each value of γ the accuracy was computed at $t = 1s$ and γ^* was selected as the value that maximized accuracy. This resulted in a specific value of γ^* for each $\mathcal{S}/\sqrt{h}$. Fig. 3B was generated using 10^6 simulations of Eq. 2 (using $\tilde{h}$ instead of h) and Eq. 7 (using $\tilde{\gamma}$ instead of γ) from $t = 0$ to $t = 1$s for a range of values of $\tilde{h}$ and $\tilde{\gamma}$. For each value of $\tilde{h}$ and $\tilde{\gamma}$, the maximal accuracy was estimated as the value of the accuracy at $t = 1$s. Fig. 3C was generated by estimating the second derivative of the curves shown in Fig. 3B for each value of $\mathcal{S}/\sqrt{h}$. Fig. 3D was generated as follows using Eq. 2 and Eq. 7. For each of the four curves, $\tilde{\gamma}$ was fixed to the value of γ^* corresponding to the reference values $\mathcal{S}/\sqrt{h} = i$ for $i = 1, 2, 3, 4$ (see Fig. 3A). For each curve, this value of $\tilde{\gamma}$ was not changed when new $\mathcal{S}/\sqrt{h}$ values were used. Then, for each curve, the maximal accuracy for the linear and nonlinear models were computed using 10^6 simulations for a range of new $\mathcal{S}/\sqrt{h}$ values. The quotient of the maximal accuracy of the linear model and the maximal accuracy of the nonlinear model is shown in the figure.

Fig. 4C-D presents the results of five hundred independent fitting procedures, performed on two different dataset sizes. The parameters for the reference dataset of trials are: $h_{\text{stim}} = 1$ Hz, $(\lambda_{\text{high}}, \lambda_{\text{low}}) = (20, 5)$ Hz, and $T = 2$ s. For each fitting procedure, the trials (either 100 or 500) were sampled uniformly without replacement from a bank of

10,000 trials. The fitting algorithm is an implementation of the Bayesian approach leading to equation (27) above. When fitting the linear model, the analytical solution from appendix D.1 was used to compute the likelihood of a single trial (P($d_k | \mathcal{T}_k, \theta$) term in Eq. (27)). When fitting the nonlinear model, Monte Carlo sampling was used instead. More specifically, the distribution of the decision variable at decision time for a given clicks stimulus, $p(y_T | \mathcal{T})$, was estimated by simulating 800 independent trajectories. Thus, each trajectory had its own independent realization of sensory noise but the realization of the stimulus (timing of the clicks) was frozen. Once the density of y_T was estimated, the likelihood term, P($d_k | \mathcal{T}_k, \theta$) in Eq. (27), could be estimated. More details on this method, such as how the number of 800 particles was chosen and how this method was validated on the linear model for which the analytical solution is available, may be found in section 3.5.5 of Radillo (2018).

In Fig 4E, up to trial number 500 on the x-axis, the same fits as in panels C-D were used to compute the relative error (y-axis). Because of the high computational cost of our fitting algorithm (Monte Carlo sampling described above), the points for 1000 trials on the x-axis were computed with only 84 independent fits per model pair (as opposed to 500 for the other points of the figure).

All panels in Fig. 5 were produced with a common dataset of 10^6 trials, generated by presenting the same sets of click streams to the evidence accumulation models. All trials had same task parameters: trial duration T = 2s; hazard rate h = 1 Hz; λ_{high} = 20 Hz and λ_{low} = 5 Hz so $S/\sqrt{h}$ = 3; and the initial state of the environment was randomly assigned with a uniform prior. For each panel of Fig. 5, we selected a pair ($m_{\text{fit}}, m_{\text{ref}}$) $\in$ {(L, L), (NL, NL)} along with a sensory noise amplitude ($\sigma \in \{0.1, 2\}$ for $\eta_i, \zeta_i \sim \mathcal{N}(\kappa, \sigma)$) to be applied to the evidence pulses from the clicks. For each possible pair of discounting parameters ((γ_1, γ_2) for linear models, (h_1, h_2) for nonlinear models), we computed the 10^6 decisions (Left or Right) and determined whether the models agreed or not. For the linear model, we used values of γ_1, γ_2 between 0 and 10, with increments of 0.1. For the nonlinear model, we used values of h_1, h_2 between 0 and 2.5, with increments of 0.1. For each decision comparison between reference model and fit model, the same click streams were used, but independent noise realizations of click perturbations were applied. The number of agreements was divided by the total number of decision comparisons to produce the color of a single point in the plot.

G | BIAS METRIC AS A FUNCTION OF SENSORY NOISE

In this section, we provide additional information about the bias in parameter recovery with the 0/1-loss function described in Section 7. Fig. 6A includes results from simulations for noise= 1 in addition to noise= 0.1 and noise= 2 also shown in Fig. 5. Bias magnitude and its dependence on sensory noise were determined as follows. Let $\theta_{\text{ref}} \in \{\gamma_1, h_1\}$ denote the discounting parameter of the reference model – this is the model used to produce the initial decision data. Let θ_{fit} denote the fit value of the discounting parameter, using 0/1-loss minimization. In Fig. 6A, θ_{ref} spans the x-axis and θ_{fit} as a function of θ_{ref} is depicted by the golden curve. After smoothing θ_{fit} with a Savitzky-Golay filter, we obtain θ_{smooth} represented by the green curves in the figure. Picking a fixed reference value for θ_{ref} (red dotted line), we then plot the bias as a function of sensory noise levels in Fig. 6B, where bias is defined as:

$$\text{bias} := \log \frac{\theta_{\text{ref}}}{\theta_{\text{smooth}}}. \tag{29}$$

The fixed values of θ_{ref} chosen were the same as in Section 6, $\gamma_1 := 6.7457, h_1 := 1$. As described in Section 7, the bias in parameter recovery with the 0/1-loss fitting procedure is more pronounced for the nonlinear model than for the linear model, and increases with sensory noise.
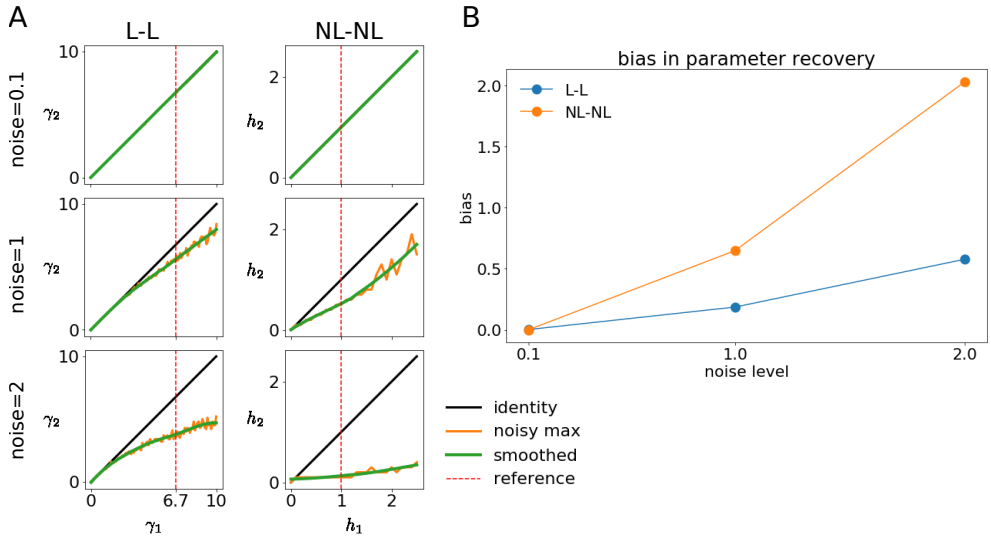
FIGURE 6 Bias in parameter recovery as a function of sensory noise. **A**: Recovered discounting parameter from the fits as a function of the reference discounting parameter used to produce the initial decision data. Top and bottom rows are reproductions of Fig. 5 while the middle row is for an intermediate level of sensory noise. The actual fit parameters (golden) were smoothed (green) in order to compute the bias in panel **B** for the reference discounting parameters indicated by the red dotted lines. The black diagonal indicates the identity line, which would correspond to perfect parameter recovery. **B:** Bias in parameter recovery Eq. (29) as a function of sensory noise, for the two model pairs (L-L in blue and NL-NL in golden).

ACKNOWLEDGEMENTS

REFERENCES

Adelson, E. H. and Bergen, J. R. (1985) Spatiotemporal energy models for the perception of motion. *JOSA A*, **2**, 284–299.

Balasubramanian, V. (1997) Statistical inference, Occam's razor, and statistical mechanics on the space of probability distributions. *Neural Comput.*, **9**, 349–368.

Ball, K. and Sekuler, R. (1982) A specific and enduring improvement in visual motion discrimination. *Science*, **218**, 697–698.

Barendregt, N. W., Josić, K. and Kilpatrick, Z. P. (2019) Analyzing dynamic decision-making models using Chapman-Kolmogorov equations. *arXiv preprint arXiv:1903.10131*.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen, M. N., Latham, P. E. and Pouget, A. (2008) Probabilistic population codes for bayesian decision making. *Neuron*, **60**, 1142–1152.

Bialek, W., Nemenman, I. and Tishby, N. (2001) Predictability, complexity, and learning. *Neural Comput.*, **13**, 2409–2463.

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.

Born, R. T. and Bradley, D. C. (2005) Structure and function of visual area mt. *Annu. Rev. Neurosci.*, **28**, 157–189.

Britten, K. H., Shadlen, M. N., Newsome, W. T. and Movshon, J. A. (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci.*, **12**, 4745–4765.

Brody, C. D. and Hanks, T. D. (2016) Neural underpinnings of the evidence accumulator. *Curr. Op. Neurobiol.*, **37**, 149–157.

Brunton, B. W., Botvinick, M. M. and Brody, C. D. (2013) Rats and humans can optimally accumulate evidence for decision-making. *Science*, **340**, 95–8.

Buesing, L., Bill, J., Nessler, B. and Maass, W. (2011) Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS Comput. Biol.*, **7**, e1002211.

Cappé, O., Moulines, E. and Rydén, T. (2005) Inference in hidden Markov models. *Springer Series in Statistics*, 652.

Dai, B., Ding, S., Wahba, G. et al. (2013) Multivariate bernoulli distribution. *Bernoulli*, **19**, 1465–1483.

Faisal, A. A., Selen, L. P. and Wolpert, D. M. (2008) Noise in the nervous system. *Nat. Rev. Neurosci.*, **9**, 292–303.

Fischer, W. and Meier-Hellstern, K. (1992) The Markov-modulated Poisson process (MMPP) cookbook. *Perform. Eval.*, **18**, 149–171.

Friedman, J., Hastie, T. and Tibshirani, R. (2001) *The elements of statistical learning*, vol. 1, chap. 7: Model Assessment and Selection. Springer series in statistics New York, NY, USA.

Friedman, J. H. (1997) On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data Min. Knowl. Discov.*, **1**, 55–77.

Gardiner, C. (2009) *Stochastic methods*, vol. 4. Springer Berlin.

Geisler, W. S. (2003) Ideal observer analysis. In *The Visual Neurosciences* (eds. L. Chalupa and J. Werner), 825–837. Boston: MIT Press.

Glaze, C. M., Filipowicz, A. L., Kable, J. W., Balasubramanian, V. and Gold, J. I. (2018) A bias–variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nat. Hum. Behav.*, **2**, 213.

Glaze, C. M., Kable, J. W. and Gold, J. I. (2015) Normative evidence accumulation in unpredictable environments. *Elife*, **4**.

Gold, J. I. and Shadlen, M. N. (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, **36**, 299–308.

— (2007) The neural basis of decision making. *Ann. Rev. Neurosci.*, **30**.

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.

Hawkins, G., Wagenmakers, E., Ratcliff, R. and Brown, S. (2015a) Discriminating evidence accumulation from urgency signals in speeded decision making. *J Neurophysiol.*, **114**, 40–47.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R. and Brown, S. D. (2015b) Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *J Neurosci.*, **35**, 2476–2484.

Huk, A. C. and Shadlen, M. N. (2005) Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J Neurosci.*, **25**, 10420–10436.

Krajbich, I. and Rangel, A. (2011) Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 13852–13857.

Langhaar, H. L. (1980) *Dimensional analysis and theory of models*. R.E. Krieger Publishing: Huntingdon, NY.

Lánský, P. (1997) Sources of periodical force in noisy integrate-and-fire models of neuronal dynamics. *Phys. Rev. E*, **55**, 2040.

Latimer, K. W., Yates, J. L., Meister, M. L., 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.

Luce, R. D. (1986) *Response times: Their role in inferring elementary mental organization*. Oxford University Press: New York, NY.

Odoemene, O., Pisupati, S., Nguyen, H. and Churchland, A. K. (2018) Visual evidence accumulation guides decision-making in unrestrained mice. *J Neurosci.*, **38**, 10143–10155.

Ohio Supercomputer Center (1987) Ohio Supercomputer Center. `http://osc.edu/ark:/19495/f5s1ph73`.

Park, H., Lueckmann, J.-M., Von Kriegstein, K., Bitzer, S. and Kiebel, S. J. (2016) Spatiotemporal dynamics of random stimuli account for trial-to-trial variability in perceptual decision making. *Sci. Rep.*, **6**, 18832.

Pesaran, M. H. and Timmermann, A. (1992) A simple nonparametric test of predictive performance. *J Business & Econ. Statistics*, **10**, 461–465.

Piet, A. T., El Hady, A. and Brody, C. D. (2018) Rats adopt the optimal timescale for evidence integration in a dynamic environment. *Nat. Commun.*, **9**, 4265.

Rabiner, L. and Juang, B. (1986) An introduction to hidden Markov models. *IEEE ASSP Mag.*, **3**, 4–16.

Radillo, A. E. (2018) *Optimal decision making in changing environments*. Ph.D. thesis, University of Houston, `https://www.math.uh.edu/~josic/content/04-members/Radillo.pdf`.

Radillo, A. E., Veliz-cuba, A., Kilpatrick, Z. P. and Josic, K. (2017) Evidence accumulation and change rate inference in dynamic environments. *Neural Comput.*, **29**, 1561–1610.

Ratcliff, R. (1978) A theory of memory retrieval. *Psychol. Rev.*, **85**, 59–108.

Ratcliff, R. and McKoon, G. (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.*, **20**, 873–922.

Ratcliff, R., Smith, P. L., Brown, S. D. and McKoon, G. (2016) Diffusion decision model: current issues and history. *Trends Cogn. Sci.*, **20**, 260–281.

Ratcliff, R. and Tuerlinckx, F. (2002) Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychon. Bull. Rev.*, **9**, 438–481.

Richardson, M. J. and Swarbrick, R. (2010) Firing-rate response of a neuron receiving excitatory and inhibitory synaptic shot noise. *Phys. Rev. Lett.*, **105**, 178102.

Roitman, J. D. and Shadlen, M. N. (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci.*, **22**, 9475–9489.

Rudemo, M. (1972) Doubly stochastic poisson processes and process control. *Adv. Appl. Probab.*, **4**, 318–338.

Shadlen, M. N., Kiani, R., Newsome, W. T., Gold, J. I., Wolpert, D. M., Zylberberg, A., Ditterich, J., de Lafuente, V., Yang, T. and Roitman, J. (2016) Comment on "Single-trial spike trains in parietal cortex reveal discrete steps during decision-making". *Science*, **351**, 1406–1406.

Shadlen, M. N. and Newsome, W. T. (1996) Motion perception: seeing and deciding. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 628–633.

Skellam, J. G. (1946) The frequency distribution of the difference between two poisson variates belonging to different populations. *J. Royal Stat. Soc. Ser. A*, **109**, 296.

Snyder, D. L. (1975) *Random point processes*. John Wiley and Sons.

Veliz-Cuba, A., Kilpatrick, Z. P. and Josić, K. (2016) Stochastic models of evidence accumulation in changing environments. *SIAM Rev.*, **58**, 264–289.

Wald, A. (1949) Note on the Consistency of the Maximum Likelihood Estimate. *Ann. Math. Stat.*, **20**, 595–601.

Wald, A. and Wolfowitz, J. (1948) Optimum character of the sequential probability ratio test. *Ann. Math. Stat.*, 326–339.

Wilson, R. C., Nassar, M. R. and Gold, J. I. (2010) Bayesian online learning of the hazard rate in change-point problems. *Neural Comput.*, **22**, 2452–2476.

— (2013) A mixture of delta-rules approximation to Bayesian inference in change-point problems. *PLoS Comput. Biol.*, **9**, e1003150.

Publication Decision 1 from Neurons, Behavior, Data analysis, and Theory on July 27, 2019

Editorial board's determination: Revise and resubmit

Comments from the editor. Sorry again for the long delay. The reviewer is more or less happy with the manuscript, there is a list of suggestions, which I would like to ask you to pay careful attention to before the paper can be accepted.

Reviewers' comments are italicized. Our responses are in plain text. Changes to the manuscript are in blue.

Comments to the author.

Summary. This is a careful and thorough paper that presents several results regarding a normative model of decision-making within dynamic environments. The specific case studied is that of an evidence accumulation task presented in Piet et al., 2018. In this task, the subject hears two streams of Poisson-timed clicks coming from their left and right sides. The click rates on both sides may switch during the trial with some hazard rate h. The subject is asked to infer which side had the higher rate of clicks when the trial ends, thus making newer information more relevant than older information. In such a task, there are four parameters: h, T, λ_{high}, λ_{low}, where h is the hazard rate, T is the duration of the trial, λ_{high} is the higher Poisson rate of clicks, and λ_{low} is the lower rate of clicks.

The authors first show that, over a broad range as long as high, low are sufficiently large, the maximal accuracy achievable by the normative model almost exactly depends on only two parameters, $\mathcal{S}/\sqrt{h}$, and hT, where $\mathcal{S} = (\lambda_{\text{high}} - \lambda_{\text{low}})/\sqrt{\lambda_{\text{high}} + \lambda_{\text{low}}}$. They use these two effective parameters to study the model. The authors first compare the behavior of the normative model to a linear model that approximates the normative model, and find that the linear model performs near-optimally if the discounting parameter is finely tuned. To further compare the models, they generated choice data from both the normative model and the linear model, and fit the choices and clicks separately to either the normative model or the linear model. They found that the parameters recovered are biased when the model used to generate data and the model used to fit the data are different. The linear model required less number of trials when it was fit to data; the discounting parameter converged to the true value used to generate data faster for the linear model than for the normative model. The authors note that different cost functions (MLE and 0/1-loss) lead to different estimates of the parameter values and one should be cautious about what cost function to use.

The authors suggest that the normative model and the linear model fits on subjects? performance at different hT and $\mathcal{S}$ provide a convenient way of identifying the decision-making strategy that the subject is employing in the dynamic clicks task.

Major comment: The authors claim that 'if an animal is extensively trained on trials with fixed parameters h and S, but subsequently interrogated using occasional trials with different task parameters,' one may be able to identify whether the subject is using the normative model or a suboptimal (or linear approximation) model. This is an important point, and it would be greatly strengthened if this claim can be supported directly with simulations and further elucidated with explanations of the specific steps that should be taken in order to identify the strategies.

Specifically, in the case where the clicks are generated with h_{stim}, and the animal performs normatively with h_{ref}, we may fit the normative model and recover h_{ref} reliably. We can compute the maximal accuracy of the normative model given the experimental $\mathcal{S}/\sqrt{h_{\text{stim}}}$, and compare it to the accuracy of the animal. However, if we fit the linear model, how should we go from here? How would one, without knowing m_{ref}, and knowing only the clicks/choice data and the experimental parameters h_{stim} and $\mathcal{S}$, identify the model that the animal is using?

Indeed, you raise important points, which we can address by taking a closer look at the accuracy ratios we have plotted in Fig. 3D. We now explain the setup of an experiment which could be run to test the sensitivity of a subject's evidence accumulation model to determine whether it is normative-like or more sensitive like the linear models we considered. Of course, in this idealized case, we are not considering internal noise, but the principles would extend to noisy models

as discussed in Piet et al (2018). In general, the linear model will not adapt well to interspersed trials with different parameters, whereas the normative model will be more robust. This is now explained in the following paragraphs we have added to the end of Section 4:

The preceding point is illustrated by the following thought experiment. Assume a subject is extensively trained on a fixed set of task parameters: $S^{\text{ref}} = 2$, $(\lambda_{\text{low}}, \lambda_{\text{high}}) = (2, 8.5)$Hz, $h = 1$Hz (peak of the red curve in Fig. 3D). We then introduce some trials with different click rates, say $S^{\text{new}} = 5$ with $(\lambda_{\text{low}}^{\text{new}}, \lambda_{\text{high}}^{\text{new}}) = (12, 53)$Hz and $h = 1$Hz, chosen so that $\kappa = \log(\lambda_{\text{high}}/\lambda_{\text{low}})$ is constant across the two conditions. We denote by $\text{Acc}_{\text{lin}}(S)$ and $\text{Acc}_{\text{norm}}(S)$ the accuracies of an observer using the linear and normative models on trials with a given S. Since the subject was trained on click rates that correspond to S^{ref}, their discounting strategy will be adapted to these values. Note that the ratio between $\text{Acc}_{\text{norm}}(S)$ and $\text{Acc}_{\text{lin}}(S)$ when $h = 1$ is the red curve in Fig. 3D. Since the ratio between $\text{Acc}_{\text{norm}}(S^{\text{ref}})$ and $\text{Acc}_{\text{lin}}(S^{\text{ref}})$ is near 1, the linear and normative models cannot be distinguished at S^{ref}. However, a subject using the normative model tuned at S^{ref}, will still perform optimally at $S \neq S^{\text{ref}}$, if κ and h are held constant. On the other hand, a linear model optimized at S^{ref}, will no longer be optimal at $S \neq S^{\text{ref}}$. This distinction is captured by the drop in the accuracy ratio along the red curve in Fig. 3D.

We can quantify the distinction between the two models by their relative difference:

$$\frac{\text{Acc}_{\text{norm}}(S^{\text{new}}) - \text{Acc}_{\text{lin}}(S^{\text{new}})}{\text{Acc}_{\text{norm}}(S^{\text{new}})} = 0.05.$$

More generally, for any decision making model, we may define the quantity

$$D_{\text{model}}(S) := \frac{\text{Acc}_{\text{norm}}(S) - \text{Acc}_{\text{model}}(S)}{\text{Acc}_{\text{norm}}(S)}$$

which will equal 0 if the model used is the normative one. If we compute $D_{\text{model}}(S)$ using responses from a real subject, one can generate curves such as those in Fig. 3D. If the curves are not constant (equal to 1), this would suggest the subject is not using an optimal model. Furthermore, a single value of S^{new} for which $D_{\text{model}}(S^{\text{new}}) \neq 0$ provides evidence that the model is not optimal.

Minor comments:

Introduction: when the term 'nonlinear model' is first used, the authors haven't yet clarified that here the term is synonymous with 'normative model.' Also, some readers might not be familiar with what a 0/1 loss function is, making that part of the Intro a little unclear for them.

It is important to note that the nonlinear model is only normative when the discounting parameter is tuned exactly, and this is why we used this specific phrasing. To clarify this, we have added the word 'nonlinear' in parentheses when mentioning the normative model in the prior sentence. We also added the following footnote: The 'nonlinear' model here refers to the family of models obtained by tuning the discounting rate away from the value defining the normative model. This detuning results in a model that is not normative.

We describe the 0/1-loss function now in the sentence following its introduction: The 0/1-loss function gives a one unit penalty on trials in which the decision predicted by the model and the data disagree, and no penalty when they agree. Therefore, minimizing this loss function leads to models that best match the trial-to-trial responses in the data rather than the response accuracy.

Page 4, last paragraph: You need the word 'alone': the phrase 'kappa does not predict the response accuracy' should probably be 'kappa alone does not predict the response accuracy.'

Thanks for pointing this out. We have modified the text as suggested.

Figure 1, caption: Might be helpful to title panel C as 'h=1' and remind readers in the caption that hT and SNR jointly determine accuracy? And for panel D, perhaps write 'Maximal accuracy of the ideal observer, at T ≫ 1?

Good suggestion. We have added the title to Fig. 1C and the sentence, "Note that hT and $\mathcal{S}/\sqrt{h}$ jointly determine accuracy." to the caption, and also the qualifier, at $T \gg 1$, to the caption of Fig. 1D.

Equation 5: Would be helpful to explicitly say here that F represents the SNR

We have added the phrase "representing the SNR" to the sentence preceding equation 5.

Paragraph immediately after equation 5: as written, it sounds a bit like F depending only on those 2 parameters follows from keeping hT fixed, although that is not what you mean.

We have changed the confusing sentence to: As indicated, $F(hT, \mathcal{S}/\sqrt{h})$ only depends on hT and $\mathcal{S}/\sqrt{h}$.

Paragraph at the end of Section 3: 'To increase the accuracy of an ideal observer, it is not enough to increase both click rates, for instance.' I didn't understand that? If everything else is kept fixed, but both lambdas grow, doesn't SNR grow and hT stay fixed?

In this situation SNR does not necessarily grow. Indeed, this depends on how the parameters are incremented. For instance, if the lambdas grow along the parabolas shown in Fig. 1D, then SNR stays constant. We have edited the sentence referenced above to explicitly state this. It now reads: To increase the accuracy of an ideal observer, it is not sufficient to increase both click rates, for instance, since the SNR stays constant if λ_{high} and λ_{low} follow the parabolas shown in Fig. 1D.

Figure 3C, label on vertical axis: why is this 'relative' accuracy?

The adjective "relative", was meant to highlight the fact that we computed the curvature of the graphs in Fig. 3B, representing functions of the relative error rather than the actual values of $\tilde{h}$ and $\tilde{\gamma}$. We agree that this was confusing, and have removed the word "relative" from the plot and added an explanation to the caption, which now says: Since the functions in panel **B** do not depend on the actual values of $\tilde{h}$ and $\tilde{\gamma}$, but rather the relative distance of these parameters from reference values, what we show in this plot are *relative* curvatures. We compare relative curvatures as $\tilde{h}$ and $\tilde{\gamma}$ do not have the same units.

The authors state that for Figure 4E, NL-L and L-NL do not converge to zero, whereas NL-NL and L-L converge to zero eventually. This is not very clear in the figure, and it would help the reader if the number of trials was larger to show this more clearly. This panel shows that for a given number of trials (in the figure, < 500 trials), the fits for the linear model were better than the normative model. How sufficient should the number of trials be for the fits to the normative model to be better than the fits to the linear model? In principle, when the normative model is fit to dataset generated by the normative model, it should eventually have lower relative error than the linear model.

We ran an additional 84 fitting procedures per model pair for training sets of 1,000 trials. We did not run more simulations as our Monte Carlo sampling method is costly (these 84 simulations took 24 hours to run on a modern 4-core laptop). We report the resulting relative errors in the new Fig. 4E. This figure now shows that the NL-NL curve does fall below the L-NL curve on average after 1000 trials. We also added the following explanation to appendix F of the revised manuscript: In Fig 4E, up to trial number 500 on the x-axis, the same fits as in panels C-D were used to compute the

3

relative error (y-axis). Because of the high computational cost of our fitting algorithm (Monte Carlo sampling described above), the points for 1000 trials on the x-axis were computed with only 84 independent fits per model pair (as opposed to 500 for the other points of the figure).

Figure 5 needs labels C and D. What was the grid of values used for h and in generating Fig. 5? There are comparisons between 0.1 and 2, but do we see a more deviating trend as a function of noise level? That is, what is the deviation like when noise is 0.5 or 1, for example? It would be great to have a summary plot with a metric of deviation separately for both the linear and normative models. What do we see when the metric is shown as a function of noise? Having this summary figure will help the authors establish the point that 'the parameters that minimize expected 0/1-loss are biased, and this bias increases with sensory noise,' and that 'the NL model exhibits this bias much more strongly than the L model.'

We have added the labels C and D to the appropriate panels in Figure 5. We have also specified the grid values used in the simulations in appendix F, using the following added passage: For the linear model, we used values of γ_1, γ_2 between 0 and 10, with increments of 0.1. For the nonlinear model, we used values of h_1, h_2 between 0 and 2.5, with increments of 0.1.

Regarding the second suggestion of measuring the bias in parameter recovery for intermediate values of noise, we have added the figure that appears below in this response letter, and is now contained in Appendix G to our manuscript. We reference this figure in the following text we have added to Section 7: See Appendix G for a possible metric of the reported bias, and its dependence on sensory noise for each model class (Fig. 6).

Appendix B that derives the SNR has typos. In the sentence right after Eq. (16), the statement is $E[N] = (\lambda_{high} - \lambda_{low})/\sqrt{(h)}$, as T approaches infinity. Should this be $E[N] = (\lambda_{high} - \lambda_{low})/h$?

Yes, we have checked Appendix B thoroughly, and corrected this typo.

Eq. (17) seems to contain an error that multiplies 2 in front of hT*exp(-hT). In the sentence right after Eq. (17), SNR=E[Delta N]/Var[Delta N]. Should this be SNR=E[Delta N]/sqrt(Var[Delta N])? In the equation that follows this sentence, h is not multiplied in the denominator.

You're right – we have now corrected these typos, and also spotted a missing factor of 2 in the denominator of the full SNR expression. The equations should all be correct now.

For convenience, one could state what S is also in the Appendix.

We have added Eq. (19) in Appendix B as a reminder of Eq. (4) for the reader.

Appendix F on the details of the simulations could be more detailed. The authors state that when fitting the nonlinear model, Monte Carlos sampling was used. Please explain further on this fitting method to the extent that the reader can replicate.

We have added the following passage to Appendix F: More specifically, the distribution of the decision variable at decision time for a given clicks stimulus, $p(y_T | \mathcal{T})$, was estimated by simulating 800 independent trajectories. Thus, each trajectory had its own independent realization of sensory noise but the realization of the stimulus (timing of the clicks) was frozen. Once the density of y_T was estimated, the likelihood term, $P(d_k | \mathcal{T}_k, \theta)$ in Eq. (27), could be estimated. More details on this method, such as how the number of 800 particles was chosen and how this method was validated on the linear model for which the analytical solution is available, may be found in section 3.5.5 of Radillo (2018).

Although the confidence intervals do not seem to contain the true parameter value, are the mismatches statistically significant in Figure 4CD? That is, are the parameter values significantly greater than the true parameter value?

We would like to note that the whiskers of each box in Fig. 4C,D do not represent confidence intervals. Instead, they represent the interval of values that are not considered outliers. More specifically, if q_1, q_3 are the first and third quartiles respectively, then the whiskers define the interval: $[q_1 - 1.5 \times (q_3 - q_1), \; q_3 + 1.5 \times (q_3 - q_1)]$.

Instead of testing the hypothesis that the MAP estimates are different than the θ_{true} value, we provide a summary statistic for the training datasets of size 500 trials. We have added the following sentences to the end of section 6: For reference datasets of size 500, 98% of the 500 MAP estimates in the L-NL fits lie strictly above γ_{true}, versus 50.4% for the corresponding L-L fits. Similarly, 86.6% of the estimates in the NL-L fits lie strictly below h_{true}, versus 44.2% for the corresponding NL-NL fits. We believe this is more informative than a p-value.

Publication Decision 2 from Neurons, Behavior, Data analysis, and Theory on August 26, 2019

Editorial board's determination: Accept

Comments from the editor. Thanks a lot for your patience with the first round of reviews. I have now checked your revision and have no further comments, which means the manuscript will be transferred to a "provisional acceptance" stage.