

Optimizing sequential decisions in the drift-diffusion model

Khanh P. Nguyen^a, Krešimir Josić^{a,b,c,f}, Zachary P. Kilpatrick^{d,e,f}

^a*Department of Mathematics, University of Houston, Houston TX 77204 (kpnguyen@math.uh.edu, josic@math.uh.edu)*

^b*Department of Biology and Biochemistry, University of Houston, Houston TX 77204*

^c*Department of BioSciences, Rice University, Houston TX 77005*

^d*Department of Applied Mathematics, University of Colorado, Boulder, Colorado 80309, USA (zpkilpat@colorado.edu).*

^e*Department of Physiology and Biophysics, University of Colorado School of Medicine, Aurora, CO 80045*

^f*equal authorship*

Abstract

To make decisions organisms often accumulate information across multiple timescales. However, most experimental and modeling studies of decision-making focus on sequences of independent trials. On the other hand, natural environments are characterized by long temporal correlations, and evidence used to make a present choice is often relevant to future decisions. To understand decision-making under these conditions we analyze how a model ideal observer accumulates evidence to freely make choices across a sequence of correlated trials. We use principles of probabilistic inference to show that an ideal observer incorporates information obtained on one trial as an initial bias on the next. This bias decreases the time, but not the accuracy of the next decision. Furthermore, in finite sequences of trials the rate of reward is maximized when the observer deliberates longer for early decisions, but responds more quickly towards the end of the sequence. Our model also explains experimentally observed patterns in decision times and choices, thus providing a mathematically principled foundation for evidence-accumulation models of sequential decisions.

Keywords: decision-making, drift-diffusion model, reward rate, sequential correlations

1. Introduction

Organismal behavior is often driven by decisions that are the result of evidence accumulated to determine the best among available options (Gold and Shadlen, 2007; Brody and Hanks, 2016). For instance, honeybee swarms use a democratic process in which each bee's opinion is communicated to the group to decide which nectar source to forage (Seeley et al., 1991). Competitive animals evaluate their opponents' attributes to decide whether to fight or flee (Stevenson and Rillich, 2012), and humans decide which stocks to buy or sell, based on individual research and social information (Moat et al., 2013). Importantly, the observations of these agents are frequently uncertain (Hsu et al., 2005; Brunton et al., 2013) so accurate decision-making requires robust evidence integration that accounts for the reliability and variety of evidence sources (Raposo et al., 2012).

The two alternative forced choice (TAFC) task paradigm has been successful in probing the behavioral trends and neural mechanisms underlying decision-making (Ratcliff, 1978). In a TAFC task subjects decide which one of two hypotheses is more likely based on noisy evidence (Gold and Shadlen, 2002; Bogacz et al., 2006). For instance, in the random dot motion discrimination task, subjects decide whether a cloud of noisy dots predominantly moves in one of two directions. Such stimuli evoke strong responses in primate motion-detecting areas, motivating their use in the experimental study of neural mechanisms underlying decision-making (Shadlen and Newsome, 2001; Gold and Shadlen, 2007). The response trends and underlying neural activity are well described by the drift-diffusion model (DDM), which associates a subject's belief with a particle drifting and diffusing between two boundaries, with decisions determined by the first boundary the particle encounters (Stone, 1960; Bogacz et al., 2006; Ratcliff and McKoon, 2008).

The DDM is popular because (a) it can be derived as the continuum limit of the statistically-optimal sequential probability ratio test (Wald and Wolfowitz, 1948; Bogacz et al., 2006); (b) it is an analytically tractable Wiener diffusion process whose summary statistics can be computed explicitly (Ratcliff and Smith, 2004; Bogacz et al.,

2006); and (c) it can be fit remarkably well to behavioral responses and neural activity in TAFC tasks with independent trials (Gold and Shadlen, 2002, 2007) (although see Latimer et al. (2015)).

However, the classical DDM does not describe many aspects of decision-making in natural environments. For instance, the DDM is typically used to model a series of independent trials where evidence accumulated during one trial is not informative about the correct choice on other trials (Ratcliff and McKoon, 2008). Organisms in nature often make a sequence of related decisions based on overlapping evidence (Chittka et al., 2009). Consider an animal deciding which way to turn while fleeing a pursuing predator: To maximize its chances of escape its decisions should depend on both its own and the predator's earlier movements (Corcoran and Conner, 2016). Animals foraging over multiple days are biased towards food sites with consistently high yields (Gordon, 1991). Thus even in a variable environment, organisms use previously gathered information to make future decisions (Dehaene and Sigman, 2012). We need to extend previous experimental designs and corresponding models to understand if and how they do so.

Even in a sequence of independent trials, previous choices influence subsequent decisions (Fernberger, 1920). Such serial response dependencies have been observed in TAFC tasks which do (Cho et al., 2002; Fründ et al., 2014) and do not (Bertelson, 1961) require accumulation of evidence across trials. For instance, a subject's response time may decrease when the current state is the same as the previous state (Pashler and Baylis, 1991). Thus, trends in response time and accuracy suggest subjects use trial history to predict the current state of the environment, albeit suboptimally (Kirby, 1976).

Goldfarb et al. (2012) examined responses in a series of dependent trials with the correct choice across trials evolving according to a two-state Markov process. The transition probabilities affected the response time and accuracy of subjects in ways well described by a DDM with biased initial conditions and thresholds. For instance, when repeated states were more likely, response times decreased in the second of two repeated trials. History-dependent biases also increase the probability of repeat responses when subjects view sequences with repetition probabilities above chance (Abrahamyan et al., 2016; Braun et al., 2018). These results suggest that an adaptive DDM with an initial condition biased towards the previous decision is a good model of human decision-making across correlated trials (Goldfarb et al., 2012).

Most earlier models proposed to explain these observations are intuitive and recapitulate behavioral data, but are at least partly heuristic (Goldfarb et al., 2012; Fründ et al., 2014). Yu and Cohen (2008) have proposed a normative model that assumes the subjects are learning non-stationary transition rates in a dependent sequence of trials. However, they did not examine the normative model for known transition rates. Such a model provides a standard for experimental subject performance, and a basis for approximate models, allowing us to better understand the heuristics subjects use to make decisions (Ratcliff and McKoon, 2008; Brunton et al., 2013).

Here, we extend previous DDMs to provide a normative model of evidence accumulation in serial trials evolving according to a two-state Markov process whose transition rate is known to the observer. We use sequential analysis to derive the posterior for the environmental states H_{\pm} given a stream of noisy observations (Bogacz et al., 2006; Veliz-Cuba et al., 2016). Under this model, ideal observers incorporate information from previous trials to bias their initial belief on subsequent trials. This decreases the average time to the next decision, but not necessarily its accuracy. Furthermore, in finite sequences of trials the rate of reward is maximized when the observer deliberates longer for early decisions, but responds more quickly towards the end of the sequence. We also show that the model agrees with experimentally observed trends in decisions.

2. A model of sequential decisions

We model a repeated TAFC task with the environmental state in each trial (H_{+} or H_{-}) chosen according to a two-state Markov process: In a sequence of n trials, the correct choices (environmental states or hypotheses) $H^{1:n} = (H^1, H^2, \dots, H^n)$ are generated so that $P(H^1 = H_{\pm}) = 1/2$ and $P(H^i = H_{\mp} | H^{i-1} = H_{\pm}) = \epsilon$ for $i = 2, \dots, n$ (See Fig. 1). When $\epsilon = 0.5$ the states of subsequent trials are independent, and ideally the decision on one trial should not bias future decisions (Ratcliff, 1978; Shadlen and Newsome, 2001; Gold and Shadlen, 2007, 2002; Bogacz et al., 2006; Ratcliff and McKoon, 2008).

When $0 \leq \epsilon < 0.5$, repetitions are more likely and trials are dependent, and evidence obtained during one trial can inform decisions on the next¹. We will show that ideal observers use their decision on the previous trial to adjust their

¹If $0.5 < \epsilon \leq 1$ states are more likely to alternate. The analysis is similar to the one we present here, so we do not discuss this case separately.

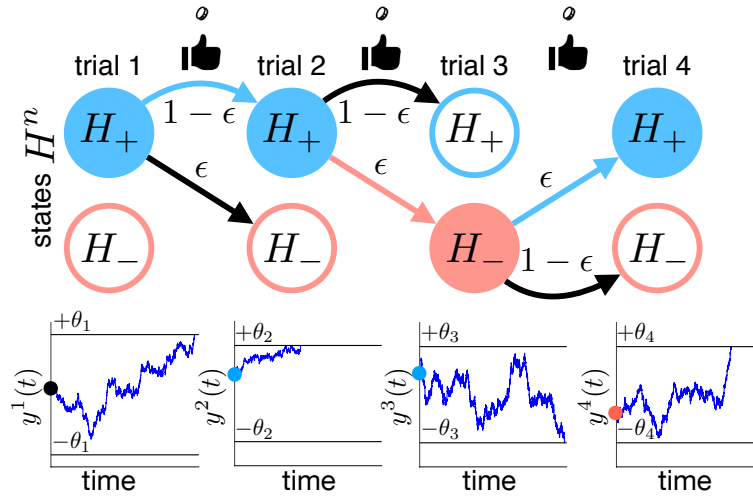


Figure 1: Sequence of TAFC trials with states H^n determined by a two-state Markov process with switching probability $\epsilon := P(H^n \neq H^{n-1})$. For instance, $H^{1:4} = (H_+, H_+, H_-, H_+)$ could stand for the drift direction of each drift-diffusion process $y^n(t)$ corresponding to the decision variable of an ideal observer making noisy measurements, $\xi^n(t)$. The decision threshold ($\pm\theta_{n-1}$) crossed in trial $n-1$ determines the sign of the initial condition in trial n : $y^n(0)$.

prior over the states at the beginning of the following trial.

Importantly, we assume that all rewards are given at the end of the trial sequence (Braun et al., 2018). Rewarding the correct choice on trial i provides unambiguous information about the state H^i , superseding the noisy information gathered over that trial. Our results can be easily extended to this case, as well as cases when rewards are only given with some probability.

Our main contribution is to derive and analyze the ideal observer model for this sequential version of the TAFC task. We use principles of probabilistic inference to derive the sequence of DDMs, and optimization to determine the decision thresholds that maximize the reward rate (RR). Due to the tractability of the DDM, the correct probability and decision times that constitute the RR can be computed analytically. We also demonstrate that response biases, such as repetitions, commonly observed in sequential decision tasks, follow naturally from this model.

3. Optimizing decisions in a sequence of uncorrelated trials

We first derive in Appendix A and summarize here the optimal evidence-accumulation model for a sequence of independent TAFC trials ($P(H^i = H^{i-1}) = 0.5$). Although these results can be found in previous work (Bogacz et al., 2006), they are crucial for the subsequent discussion, and we thus provide them for completeness. *A reader familiar with these classical results can skip to Section 4, and refer back to results in this section as needed.*

The drift-diffusion model (DDM) can be derived as the continuum limit of a recursive equation for the log-likelihood ratio (LLR) (Wald and Wolfowitz, 1948; Gold and Shadlen, 2002; Bogacz et al., 2006). When the environmental states are uncorrelated and unbiased, an ideal observer has no bias at the start of each trial.

3.1. Drift-diffusion model for a single trial

We assume that on a trial the observer integrates a stream of noisy measurements of the true state, H^1 . If these measurements are conditionally independent, the functional central limit theorem yields the DDM for the scaled LLR, $y^1(t) = D \log \frac{P(H^1=H_+|\text{observations})}{P(H^1=H_-|\text{observations})}$, after observation time t ,

$$dy^1 = g^1 dt + \sqrt{2D}dW. \quad (1)$$

Here W is a Wiener process, the drift $g^1 \in g_{\pm}$ depends on the environmental state, and D is the variance which depends on the noisiness of each observation.

For simplicity and consistency with typical random dot kinetogram tasks (Schall, 2001; Gold and Shadlen, 2007), we assume each drift direction is of equal unit strength: $g_+ = -g_- = 1$, and task difficulty is controlled by scaling the variance² through D . The initial condition is determined by the observer's prior bias, $y^1(0) = D \ln[P(H^1 = H_+)/P(H^1 = H_-)]$. Thus for an unbiased observer, $y^1(0) = 0$.

There are two primary ways of obtaining a response from the DDM given by Eq. (1) that mirror common experimental protocols (Shadlen and Newsome, 2001; Gold and Shadlen, 2002, 2007): An ideal observer interrogated at a set time, $t = T$ responds with $\text{sign}(y(T)) = \pm 1$ indicating the more likely of the two states, $H^1 = H_{\pm}$, given the accumulated evidence. On the other hand, an observer free to choose their response time can trade speed for accuracy in making a decision. This is typically modeled in the DDM by defining a decision threshold, θ_1 , and assuming that at the first time, T_1 , at which $|y^1(T_1)| \geq \theta_1$, the evidence accumulation process terminates, and the observer chooses H_{\pm} if $\text{sign}(y^1(T_1)) = \pm 1$.

The probability, c_1 , of making a correct choice in the free response paradigm can be obtained using the Fokker-Planck (FP) equation corresponding to Eq. (1) (Gardiner, 2009). Given the initial condition $y^1(0) = y_0^1 = D \ln[P(H^1 = H_+)/P(H^1 = H_-)]$, threshold θ_1 , and state $H^1 = H_+$, the probability of an exit through either boundary $\pm\theta_1$ is

$$\pi_{\theta_1}(y_0^1) = \frac{1 - e^{-(y_0^1 + \theta_1)/D}}{1 - e^{-2\theta_1/D}}, \quad \pi_{-\theta_1}(y_0^1) = \frac{e^{-(y_0^1 + \theta_1)/D} - e^{-2\theta_1/D}}{1 - e^{-2\theta_1/D}}, \quad (2)$$

simplifying at $y_0^1 = 0$ to

$$\pi_{\theta_1}(0) = \frac{1}{1 + e^{-\theta_1/D}} =: c_1, \quad \pi_{-\theta_1}(0) = \frac{e^{-\theta_1/D}}{1 + e^{-\theta_1/D}} = 1 - c_1. \quad (3)$$

An exit through the threshold θ_1 results in a correct choice of $H^1 = H_+$, so $c_1 = \pi_{\theta_1}(0)$. The correct probability c_1 increases with θ_1 , since more evidence is required to reach a larger threshold. Defining the decision in trial 1 as $d_1 = \pm 1$ if $y^1(T_1) = \pm\theta_1$, Bayes' rule implies

$$\frac{1}{1 + e^{-\theta_1/D}} = c_1 = P(d_1 = \pm 1 | H^1 = H_{\pm}) = P(H^1 = H_{\pm} | d_1 = \pm 1), \quad (4a)$$

$$\frac{e^{-\theta_1/D}}{1 + e^{-\theta_1/D}} = 1 - c_1 = P(d_1 = \pm 1 | H^1 = H_{\mp}) = P(H^1 = H_{\mp} | d_1 = \pm 1), \quad (4b)$$

since $P(H^1 = H_{\pm}) = P(d_1 = \pm 1) = 1/2$. Rearranging the expressions in Eq. (4) and isolating θ_1 relates the threshold θ_1 to the LLR given a decision $d_1 = \pm 1$:

$$\pm\theta_1 = D \ln \frac{P(H^1 = H_+ | d_1 = \pm 1)}{P(H^1 = H_- | d_1 = \pm 1)}. \quad (5)$$

3.2. Tuning performance via speed-accuracy tradeoff

Increasing θ_1 increases the probability of a correct decision, and the average time to make a decision, DT_1 . Humans and other animals balance speed and accuracy to maximize the rate of correct decisions (Chittka et al., 2009; Bogacz et al., 2010). This is typically quantified using the reward rate (RR) (Gold and Shadlen, 2002):

$$\text{RR}_1(\theta_1) = \frac{c_1}{DT_1 + T_D}, \quad (6)$$

where c_1 is the probability of a correct decision, DT_1 is the mean time required for $y^1(t)$ to reach either threshold $\pm\theta_1$, and T_D is the prescribed time delay to the start of the next trial (Gold and Shadlen, 2002; Bogacz et al., 2006). Eq. (6) increases with c_1 (accuracy) and with inverse time $1/(DT_1 + T_D)$ (speed). This usually leads to a nonmonotonic dependence of RR_1 on the threshold, θ_1 , since increasing θ_1 increases accuracy, but decreases speed (Gold and Shadlen, 2002; Bogacz et al., 2006).

²An arbitrary drift amplitude g can also be scaled out via a change of variables $y = g\tilde{y}$, so the resulting DDM for \tilde{y} has unit drift, and $\tilde{D} = D/g^2$.

The average response time, DT_1 , can be obtained as the solution of a mean exit time problem for the Fokker-Planck (FP) equation for $p(y^1, t)$ with absorbing boundaries, $p(\pm\theta_1, t) = 0$ (Bogacz et al., 2006; Gardiner, 2009). Since the RR is determined by the average decision time over all trials, we compute the unconditional mean exit time,

$$T(y_0^1; \theta_1) = \theta_1 \left[\frac{e^{\theta_1/D} + e^{-\theta_1/D} - 2e^{-y_0^1/D}}{e^{\theta_1/D} - e^{-\theta_1/D}} \right] - y_0^1, \quad (7)$$

which for $y_0^1 = 0$ simplifies to

$$T(0; \theta_1) = \theta_1 \left[\frac{1 - e^{-\theta_1/D}}{1 + e^{-\theta_1/D}} \right] = DT_1. \quad (8)$$

Plugging this expression into the RR as defined in Eq. (6), assuming $y_0^1 = 0$, we have

$$\text{RR}_1(\theta_1) = \left(\theta_1 \left[1 - e^{-\theta_1/D} \right] + T_D \left[1 + e^{-\theta_1/D} \right] \right)^{-1}.$$

We can identify the maximum of $\text{RR}_1(\theta_1) > 0$ by finding the minimum of its reciprocal, $1/\text{RR}_1(\theta_1)$ (Bogacz et al., 2006),

$$\theta_1^{\text{opt}} = T_D + D - D\mathcal{W}(\exp[(T_D + D)/D]). \quad (9)$$

Here $\mathcal{W}(f)$ is the Lambert W function (the inverse of $f(\mathcal{W}) = \mathcal{W}e^{\mathcal{W}}$). In the limit $T_D \rightarrow 0$, we have $\theta_1^{\text{opt}} \rightarrow 0$, and Eq. (9) defines nontrivial optimal thresholds at $T_D > 0$.

Having established a policy for optimizing performance in a sequence of independent TAFC trials, in the next section we consider dependent trials. We can again explicitly compute the RR function for sequential TAFC trials with states, $H^{1:n} = (H^1, H^2, H^3, \dots, H^n)$, evolving according to a two-state Markov process with parameter $\epsilon := \text{P}(H^{n+1} \neq H^n)$. Unlike above, where $\epsilon = 0.5$, we will see that when $\epsilon \in [0, 0.5)$, an ideal observer starts each trial H^n for $n \geq 2$ by using information obtained over the previous trial, to bias their initial belief, $y_0^n \neq 0$.

4. Integrating information across two correlated trials

We first focus on the case of two sequential TAFC trials. Both states are equally likely at the first trial, $\text{P}(H^1 = H_{\pm}) = 1/2$, but the state at the second trial can depend on the state of the first, $\epsilon := \text{P}(H^2 = H_{\mp} | H^1 = H_{\pm})$. On both trials an observer makes observations, ξ_s^n , with conditional density $f_{\pm}(\xi)$ if $H^n = H_{\pm}$ to infer the state H^n . The observer also uses information from trial 1 to infer the state at trial 2. All information about H^1 can be obtained from the decision variable, $d_1 = \pm 1$, and ideal observers use the decision variable to set their initial belief, $y_0^2 \neq 0$, at the start of trial 2. We later show that the results for two trials can be extended to trial sequences of arbitrary length with states generated according to the same two-state Markov process.

4.1. Optimal observer model for two correlated trials

The first trial is equivalent to the single trial case discussed in section 3.1. Assuming each choice is equally likely, $\text{P}(H^1 = H_{\pm}) = 1/2$, no prior information exists. Therefore, the decision variable, $y^1(t)$, satisfies the DDM given by Eq. (1) with $y^1(0) = 0$. This generates a decision $d_1 \in \pm 1$ if $y^1(T_1) = \pm\theta_1$, so that the probability of a correct decision c_1 is related to the threshold θ_1 by Eq. (4). Furthermore, since $\epsilon = \text{P}(H^2 = H_{\mp} | H^1 = H_{\pm})$, it follows that $1 - \epsilon = \text{P}(H^2 = H_{\pm} | H^1 = H_{\pm})$, which means

$$\begin{aligned} \text{P}(H^2 = H_{\pm} | d_1 = \pm 1) &= \text{P}(H^2 = H_{\pm} | H^1 = H_{\pm}) \text{P}(H^1 = H_{\pm} | d_1 = \pm 1) \\ &\quad + \text{P}(H^2 = H_{\mp} | H^1 = H_{\mp}) \text{P}(H^1 = H_{\mp} | d_1 = \pm 1) \\ &= (1 - \epsilon)c_1 + \epsilon(1 - c_1), \end{aligned} \quad (10)$$

and, similarly,

$$\text{P}(H^2 = H_{\mp} | d_1 = \pm 1) = (1 - \epsilon)(1 - c_1) + \epsilon c_1. \quad (11)$$

As the decision, $d_1 = \pm 1$, determines the probability of each state at the end of trial 1, individual observations, $\xi_{1:s}^1$, are not needed to define the belief of the ideal observer at the outset of trial 2. The ideal observer uses a sequence of observations, $\xi_{1:s}^2$, and their decision on the previous trial, d_1 , to arrive at the probability ratio

$$R_s^2 = \frac{P(H^2 = H_+ | \xi_{1:s}^2, d_1)}{P(H^2 = H_- | \xi_{1:s}^2, d_1)} = \frac{P(\xi_{1:s}^2 | H^2 = H_+) P(H^2 = H_+ | d_1)}{P(\xi_{1:s}^2 | H^2 = H_-) P(H^2 = H_- | d_1)}. \quad (12)$$

Taking the logarithm, and applying conditional independence of the measurements $\xi_{1:s}^2$, we have that

$$L_s^2 = \ln R_s^2 = \sum_{j=1}^s \ln \frac{f_+(\xi_j^2)}{f_-(\xi_j^2)} + \ln \frac{P(H^2 = H_+ | d_1)}{P(H^2 = H_- | d_1)},$$

indicating that $L_0^2 = \ln [P(H^2 = H_+ | d_1) / P(H^2 = H_- | d_1)]$. Taking the temporal continuum limit as in Appendix A, we find

$$dy^2 = g^2 dt + \sqrt{2D} dW, \quad (13)$$

with the Wiener process, W , and variance defined as in Eq. (1). The drift $g^2 \in \pm 1$ is determined by $H^2 = H_{\pm}$.

Furthermore, the initial belief is biased in the direction of the previous decision, as the observer knows that states are correlated across trials. The initial condition for Eq. (13) is therefore

$$y^2(0; d_1 = \pm 1) = D \ln \frac{P(H^2 = H_+ | d_1 = \pm 1)}{P(H^2 = H_- | d_1 = \pm 1)} = \pm D \ln \frac{(1 - \epsilon)e^{\theta_1/D} + \epsilon}{\epsilon e^{\theta_1/D} + (1 - \epsilon)} = \pm y_0^2, \quad (14)$$

where we have used Eqs. (4), (10), and (11). Note that $y_0^2 \rightarrow 0$ in the limit $\epsilon \rightarrow 0.5$, so no information is carried forward when states are uncorrelated across trials. In the limit $\epsilon \rightarrow 0$, we find $y_0^2 \rightarrow \theta_1$, so the ending value of the decision variable $y^1(T_1) = \pm \theta_1$ in trial 1 is carried forward to trial 2, since there is no change in environmental state from trial 1 to 2. For $\epsilon \in (0, 1/2)$, the information gathered on the previous trial provides partial information about the next state, and we have $y_0^2 \in (0, \theta_1)$.

We assume that the decision variable, $y^2(t)$, evolves according to Eq. (13), until it reaches a threshold $\pm \theta_2$, at which point the decision $d_2 = \pm 1$ is registered. The full model is thus specified by,

$$dy^j = g^j dt + \sqrt{2D} dW, \quad \text{until } |y^j(T_j)| \geq \theta_j \mapsto d_j = \text{sign}(y^j(T_j)), \quad (15a)$$

$$y^1(0) = 0, \quad y^2(0) = \pm y_0^2 \text{ if } d_1 = \pm 1, \quad \text{and } g^j = \pm 1 \text{ if } H^j = H_{\pm}, j = 1, 2. \quad (15b)$$

The above analysis is easily extended to the case of arbitrary $n \geq 2$, but before we do so, we analyze the impact of state correlations on the reward rate (RR) and identify the decision threshold values $\theta_{1:2}$ that maximize the RR.

Remark: As noted earlier, we assume the total reward is given at the end of the experiment, and not after each trial (Braun et al., 2018). A reward for a correct choice provides the observer with complete information about the state H^n , so that

$$y^2(0; H^1 = \pm 1) = D \ln \frac{P(H^2 = H_+ | H^1 = \pm 1)}{P(H^2 = H_- | H^1 = \pm 1)} = \pm D \ln \frac{(1 - \epsilon)}{\epsilon}.$$

A similar result holds in the case that the reward is only given with some probability.

4.2. Performance for constant decision thresholds

We next show how varying the decision thresholds impacts the performance of the observer. For simplicity, we first assume the same threshold is used in both trials, $\theta_{1:2} = \theta$. We quantify performance using a reward rate (RR) function across both trials:

$$\text{RR}_{1:2} = \frac{c_1 + c_2}{DT_1 + DT_2 + 2T_D}, \quad (16)$$

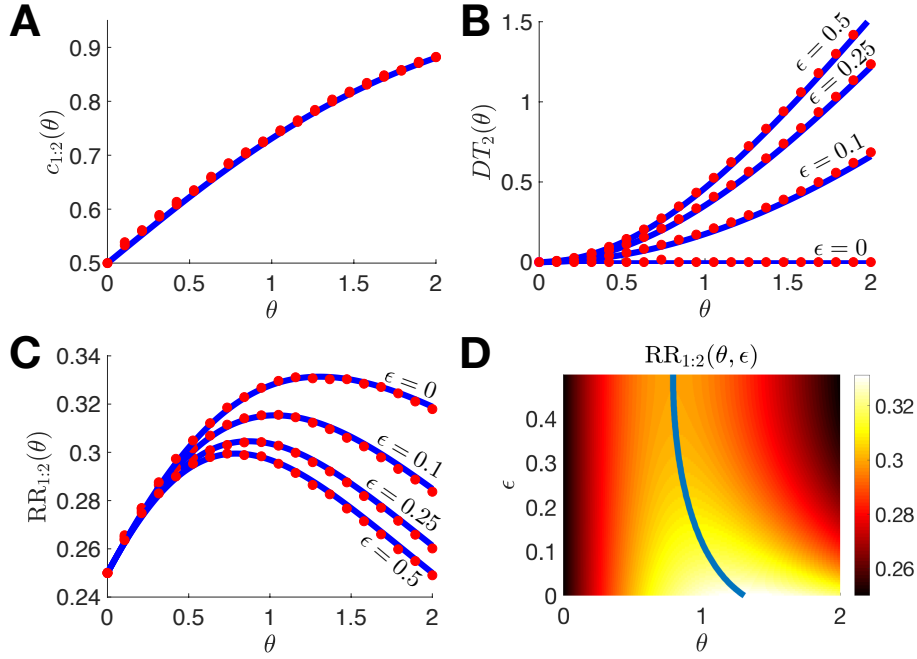


Figure 2: The reward rate (RR) depends on the state change rate $\epsilon := P(H^2 \neq H^1)$ for a constant threshold $\theta := \theta_{1,2}$. Here and below stochastic simulations (red dots) are compared to theoretical predictions (solid blue). **A.** The probability of a correct response in trials 1 and 2 ($c_{1,2}$) only depends on the threshold θ , and increases with θ . **B.** Decision time DT_2 in trial 2 increases with the threshold θ and with the change rate, ϵ . Threshold crossings happen sooner on average for lower ϵ , since trial 2 is initiated closer to threshold. **C.** The reward rate (RR) as defined by Eq. (16) is unimodal. **D.** Colormap plot of RR as a function of θ and ϵ . The maximal RR increases and occurs at higher thresholds, θ , when state changes are less likely (lower ϵ): As ϵ decreases, the observer can afford to integrate evidence for a longer period of time in trial 1, since trial 2 will be much shorter (as shown in **B**). Here and in subsequent figures we used delay time $T_D = 2$ and noise amplitude $D = 1$.

where c_j is the probability of correct responses on trial $j = 1, 2$, DT_j is the mean decision time in trial j , and T_D is the time delay after each trial. The expected time until the reward is thus $DT_1 + DT_2 + 2T_D$. The reward rate will increase as the correct probability in each trial increases, and as decision time decreases. However, the decision time in trial 2 will depend on the outcome of trial 1.

Similar to our finding for trial 1, the decision variable $y^2(t)$ in trial 2 determines $LLR_2(t)$, since $y^2 = D \cdot LLR_2$. Moreover, the probability of a correct responses on each trial, c_1 and c_2 are *equal*, and determined by the threshold, $\theta = \theta_{1,2}$, and noise amplitude, D (See Appendix B).

The average time until a decision in trial 2 is (See Appendix C)

$$DT_2 = \frac{1 - e^{-\theta/D}}{1 + e^{-\theta/D}} \left[\theta - (1 - 2\epsilon)D \ln \frac{(1 - \epsilon)e^{\theta/D} + \epsilon}{\epsilon e^{\theta/D} + (1 - \epsilon)} \right]. \quad (17)$$

Notice, in the limit $\epsilon \rightarrow 0.5$, Eq. (17) reduces to Eq. (8) for $\theta_1 = \theta$ as expected. Furthermore, in the limit $\epsilon \rightarrow 0$, $DT_2 \rightarrow 0$, since decisions are made immediately in an unchanging environment as $y_0^2(0) = \pm\theta$.

We can now use the fact that $c_1 = c_2$ to combine Eqs. (4), and (17) with Eq. (16) to find

$$RR_{1,2}(\theta) = 2 \left(\left[1 - e^{-\theta/D} \right] \left[2\theta - (1 - 2\epsilon)D \ln \frac{(1 - \epsilon)e^{\theta_1/D} + \epsilon}{\epsilon e^{\theta_1/D} + (1 - \epsilon)} \right] + 2T_D \left[1 + e^{-\theta/D} \right] \right)^{-1},$$

which can be maximized using numerical optimization.

Our analysis shows that the correct probability in both trials, $c_{1,2}$, increases with θ and does not depend on the transition probability ϵ (Fig. 2A). In addition, the decision time in trial 2, DT_2 , increases with θ and ϵ (Fig. 2B). At smaller transition rates, ϵ , more information is carried forward to the second trial, so the decision variable y_0^2 is, on

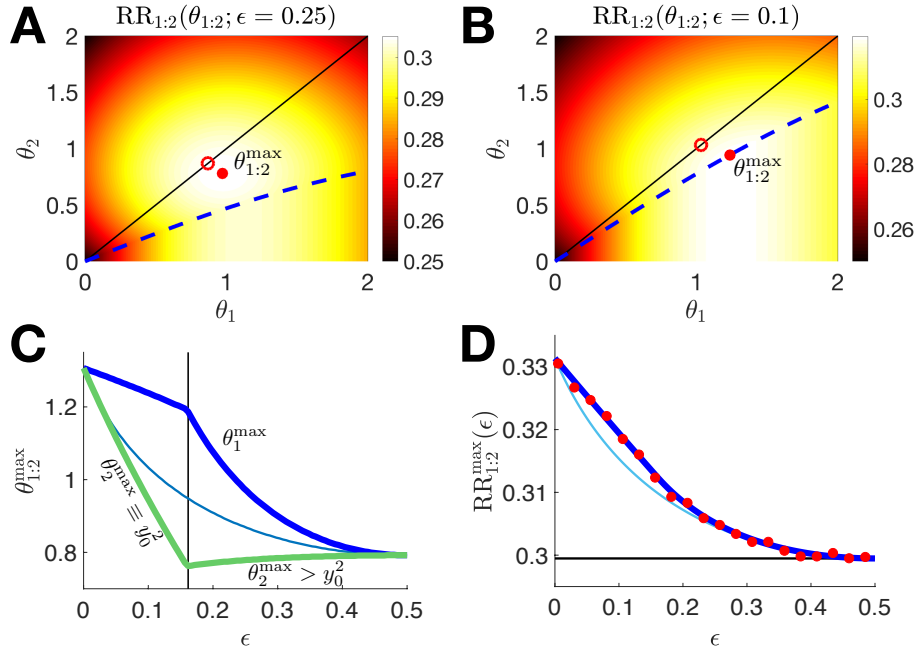


Figure 3: **A,B.** Reward rate (RR), given by Eq. (20), depends on the change rate ϵ and thresholds $\theta_{1,2}$. For $\epsilon = 0.25$ (panel **A**), the pair $\theta_{1,2}^{\max}$ that maximizes RR (red dot) satisfies $\theta_1^{\max} > \theta_2^{\max}$. Compare to the pair $\theta_{1,2}^{\max}$ (red circle) for $\theta = \theta_{1,2}$ as in Fig. 2D. The boundary $\theta_2 = y_0^2(\theta_1)$, obtained from Eq. (14), is shown as a dashed blue line. For $\epsilon = 0.1$ (panel **B**), the optimal choice of thresholds (red dot) is on the boundary $\theta_2 = y_0^2(\theta_1)$. **C.** Decreasing ϵ from 0.5, initially results in an increase in θ_1^{\max} , and decrease in θ_2^{\max} . At sufficiently small value of ϵ , the second threshold collides with the initial condition, $\theta_2 = y_0^2(\theta_1)$ (vertical black line). After this point, both θ_1^{\max} and θ_2^{\max} increase with ϵ , and meet at $\epsilon = 0$. The optimal constant threshold, θ^{\max} , is represented by a thin blue line. **D.** $RR_{1,2}^{\max}(\epsilon)$ decreases as ϵ is increased, since the observer carries less information from trial 1 to trial 2. The thin blue line shows $RR_{1,2}^{\max}$ for the constant threshold case.

average, closer to the threshold θ . This shortens the average decision time, and increases the RR (Fig. 2C, D). Lastly, note that the threshold value, θ , that maximizes the RR is higher for lower values of ϵ , as the observer can afford to require more evidence for a decision when starting trial 2 with more information.

4.3. Performance for dynamic decision thresholds

We next ask whether the RR given by Eq. (16) can be increased by allowing unequal decision thresholds between the trials, $\theta_1 \neq \theta_2$. As before, the probability of a correct response, $c_1 = \pi_{\theta_1}(0)$, and mean decision time, DT_1 , in trial 1 are given by Eqs. (3) and (8). The correct probability c_2 in trial 2 is again determined by the decision threshold, and noise variance, D , as long as $\theta_2 > y_0^2$. However, when $\theta_2 < y_0^2$, the response in trial 2 is instantaneous ($DT_2 = 0$) and $c_2 = (1 - \epsilon)c_1 + \epsilon(1 - c_1)$. Therefore, the probability of a correct response on trial 2 is defined piecewise

$$c_2 = \begin{cases} \frac{1}{1 + e^{-\theta_2/D}} & : \theta_2 > y_0^2, \\ \frac{(1 - \epsilon) + \epsilon e^{-\theta_1/D}}{1 + e^{-\theta_1/D}} & : \theta_2 \leq y_0^2. \end{cases} \quad (18)$$

We will show in the next section that this result extends to an arbitrary number of trials, with an arbitrary sequence of thresholds, $\theta_{1,n}$.

In the case $\theta_2 > y_0^2$, the average time until a decision in trial 2, is (See Appendix C)

$$DT_2 = \theta_2 \frac{1 - e^{-\theta_2/D}}{1 + e^{-\theta_2/D}} - \frac{(1 - 2\epsilon)(1 - e^{-\theta_1/D})}{1 + e^{-\theta_1/D}} D \ln \frac{(1 - \epsilon)e^{\theta_1/D} + \epsilon}{\epsilon e^{\theta_1/D} + (1 - \epsilon)}, \quad (19)$$

which allows us to compute the reward rate function $RR_{1:2}(\epsilon) = RR_{1:2}(\theta_{1:2}; \epsilon)$:

$$RR_{1:2}(\epsilon) = \begin{cases} \frac{E_+^1 + E_+^2}{E_+^2 E_-^1 [\theta_1 - (1 - 2\epsilon)y_0^2] + E_+^1 [\theta_2 E_-^2 + 2T_D E_+^2]} & : \theta_2 > y_0^2, \\ \frac{2(1 - \epsilon) + \epsilon E_+^1}{\theta_1 E_-^1 + 2T_D E_+^1} & : \theta_2 \leq y_0^2, \end{cases} \quad (20)$$

where $E_{\pm}^j = 1 \pm e^{-\theta_j/D}$. This reward rate is convex as a function of the thresholds, $\theta_{1:2}$, for all examples we examined.

The pair of thresholds, $\theta_{1:2}^{\max}$, that maximize RR satisfy $\theta_1^{\max} \geq \theta_2^{\max}$, so that the optimal observer typically decreases their decision threshold from trial 1 to trial 2 (Fig. 3A,B). This shortens the mean time to a decision in trial 2, since the initial belief y_0^2 is closer to the decision threshold more likely to be correct. As the change probability ϵ decreases from 0.5, initially the optimal θ_2^{\max} decreases and θ_1^{\max} increases (Fig. 3C), and θ_2^{\max} eventually colides with the boundary $y_0^2 = \theta_2$. For smaller values of ϵ , RR is thus maximized when the observer accumulates information on trial 1, and then makes the same decision instantaneously on trial 2, so that $c_2 = (1 - \epsilon)c_1 + \epsilon(1 - c_1)$ and $DT_2 = 0$. As in the case of a fixed threshold, the maximal reward rate, $RR_{1:2}^{\max}(\epsilon)$, increases as ϵ is decreased (Fig. 3D), since the observer starts trial 2 with more information. Surprisingly, despite different strategies, there is no big increase in $RR_{1:2}^{\max}$ compared to constant thresholds, with the largest gain at low values of ϵ .

Remark: In the results displayed for our RR maximization analysis (Fig. 2 and 3), we have fixed $T_D = 2$ and $D = 1$. We have also examined these trends in the RR at lower and higher values of T_D and D (plots not shown), and the results are qualitatively similar. The impact of dynamic thresholds $\theta_{1:2}$ on the RR are slightly more pronounced when T_D is small or when D is large.

5. Multiple correlated trials

To understand optimal decisions across multiple correlated trials, we again assume that the states evolve according to a two-state Markov Chain with $\epsilon := P(H^{j+1} = H_{\pm} | H^j = H_{\pm})$. Information gathered on one trial again determines the initial belief on the next. We first discuss optimizing the RR for a constant threshold for a fixed number n of trials known to the observer. When decision thresholds are allowed to vary between trials, optimal thresholds typically decrease across trials, $\theta_j^{\max} \leq \theta_{j-1}^{\max}$.

5.1. Optimal observer model for multiple trials

As in the case of two trials, we must consider the caveats associated with instantaneous decisions in order to define initial conditions y_0^j , probabilities of correct responses c_j , and mean decision times DT_j . The same argument leading to Eq. (13) shows that the decision variable evolves according to a linear drift diffusion process on each trial. The probability of a correct choice is again defined by the threshold, $c_j = 1/(1 + e^{-\theta_j/D})$ when $\theta_j > y_0^j$, and $c_j = (1 - \epsilon)c_{j-1} + \epsilon(1 - c_{j-1})$ when $\theta_j \leq y_0^j$, as in Eq. (18). Therefore, c_j may be defined iteratively for $j \geq 2$, as

$$c_j = \begin{cases} \frac{1}{1 + e^{-\theta_j/D}} & : \theta_j > y_0^j, \\ (1 - \epsilon)c_{j-1} + \epsilon(1 - c_{j-1}) & : \theta_j \leq y_0^j, \end{cases} \quad (21)$$

with $c_1 = 1/(1 + e^{-\theta_1/D})$ since $y_0^1 = 0$.

The probability c_j quantifies the belief at the end of trial j . Hence, as in Eq. (14), the initial belief of an ideal observer on trial $j + 1$ has the form

$$y^{j+1}(0; d_j = \pm 1) = \pm D \ln \frac{(1 - \epsilon)c_j + \epsilon(1 - c_j)}{\epsilon c_j + (1 - \epsilon)(1 - c_j)} = \pm y_0^{j+1}, \quad (22)$$

and the decision variable $y^j(t)$ within trials obeys a DDM with threshold θ_j as in Eq. (13). A decision $d_j = \pm 1$ is then registered when $y^j(T_j) = \pm \theta_j$, and if $\theta_j \leq y_0^j$, then $d_j = d_{j-1}$ and $T_j = 0$.

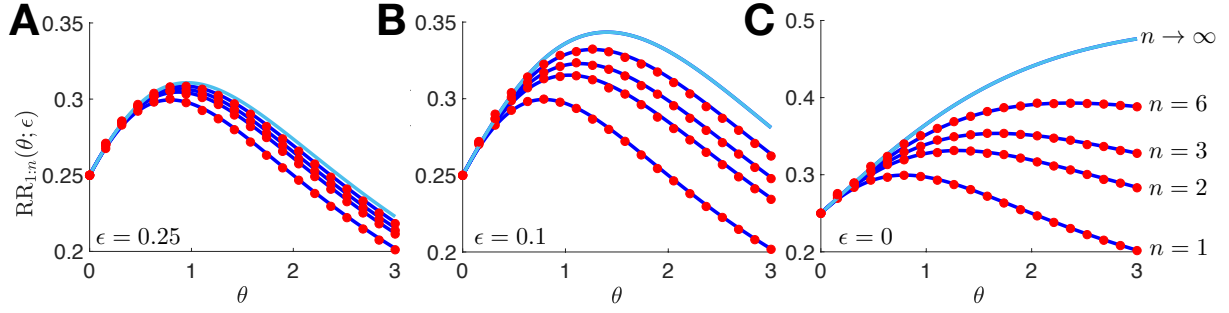


Figure 4: Reward rate ($RR_{1:n}$) defined by Eq. (24) increases with n , and as ϵ is decreased, as shown for $\epsilon = 0.25$ (panel A); $\epsilon = 0.1$ (panel B); and $\epsilon = 0$ (panel C). The optimal decision threshold, θ^{\max} , increases with n . The limiting reward rates $RR_{1:n}(\theta; \epsilon) \rightarrow RR_{\infty}(\theta; \epsilon)$ as $n \rightarrow \infty$ defined by Eq. (25) are shown in light blue.

5.2. Performance for constant decision thresholds

With n trials, the RR is the sum of correct probabilities divided by the total decision and delay time:

$$RR_{1:n} = \frac{\sum_{j=1}^n c_j}{\sum_{j=1}^n DT_j + nT_D}. \quad (23)$$

The probability of a correct choice, $c_j = 1/(1 + e^{-\theta/D})$, is constant across trials, and determined by the fixed threshold θ . It follows that $y_0^2 = y_0^3 = \dots = y_0^n = y_0$. Eq. (22) then implies that $\theta > y_0^j$ for $\epsilon > 0$ and $y_0^j = \theta$ if $\epsilon = 0$.

As y_0^j is constant on the 2nd trial and beyond, the analysis of the two trial case implies that the mean decision time is given by an expression equivalent to Eq. (17) for $j \geq 2$. Using these expressions for c_j and DT_j in Eq. (23), we find

$$RR_{1:n}(\theta; \epsilon) = n \left(\left[1 - e^{-\theta/D} \right] \left[n\theta - (n-1)(1-2\epsilon)y_0 \right] + nT_D \left[1 + e^{-\theta/D} \right] \right)^{-1}. \quad (24)$$

As n is increased, the threshold value, θ^{\max} , that maximizes the RR increases across the full range of ϵ values (Fig. 4): As the number of trials increases, trial 1 has less impact on the RR. Longer decision times on trial 1 impact the RR less, allowing for longer accumulation times (higher thresholds) in later trials. In the limit of an infinite number of trials ($n \rightarrow \infty$), Eq. (24) simplifies to

$$RR_{\infty}(\theta; \epsilon) = \left(\left[1 - e^{-\theta/D} \right] \left[\theta - (1-2\epsilon)y_0 \right] + T_D \left[1 + e^{-\theta/D} \right] \right)^{-1}, \quad (25)$$

and we find θ^{\max} of $RR_{\infty}(\theta; \epsilon)$ sets an upper limit on the optimal θ for $n < \infty$.

Interestingly, in a static environment ($\epsilon \rightarrow 0$),

$$\lim_{\epsilon \rightarrow 0} RR_{\infty}(\theta; \epsilon) = \frac{1}{T_D \left[1 + e^{-\theta/D} \right]}, \quad (26)$$

so the decision threshold value that maximizes RR diverges, $\theta^{\max} \rightarrow \infty$ (Fig. 4C). Intuitively, when there are many trials ($n \gg 1$), the price of a long wait for a high accuracy decision in the first trial ($c_1 \approx 1$ but $DT_1 \gg 1$ for $\theta \gg 1$) is offset by the reward from a large number ($n-1 \gg 1$) of subsequent, instantaneous, high accuracy decisions ($c_j \approx 1$ but $DT_2 = 0$ for $\theta \gg 1$).

5.3. Dynamic decision thresholds

In the case of dynamic thresholds, the RR function, Eq. (23), can be maximized for an arbitrary number of trials, n . The probability of a correct decision, c_j , is given by the more general, iterative version of Eq. (21). Therefore, while analytical results can be obtained for constant thresholds, numerical methods are needed to find the sequence of optimal dynamic thresholds.

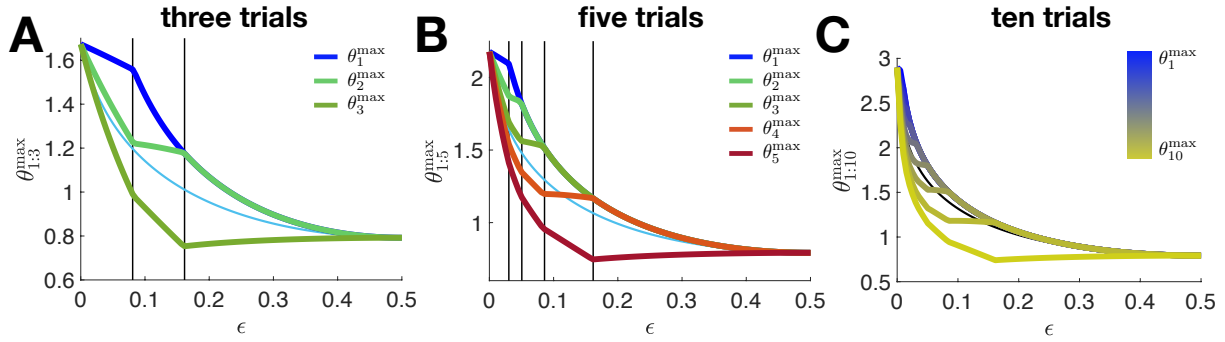


Figure 5: Optimal sets of thresholds, $\theta_{1:n}^{\max}$, for n trials with correlated states. **A.** For $n = 3$, the set of optimal thresholds $\theta_{1:3}^{\max}$ obeys the ordering $\theta_1^{\max} \geq \theta_2^{\max} \geq \theta_3^{\max}$ for all values of ϵ , converging as $\epsilon \rightarrow 0.5^-$ and as $\epsilon \rightarrow 0^+$. Vertical lines denote points below which $\theta_j^{\max} = y_0^j$ successively for $j = 3, 2, 1$. **B, C.** Ordering of optimal thresholds is consistent in the case of $n = 5$ (panel **B**) and $n = 10$ trials (panel **C**).

The expression for the mean decision time, DT_j , on trial j is determined by marginalizing over whether or not the initial state y_0^j aligns with the true state H^j , yielding

$$DT_j = \begin{cases} \theta_j \frac{1 - e^{-\theta_j/D}}{1 + e^{-\theta_j/D}} - (1 - 2\epsilon)(2c_{j-1} - 1)y_0^j & : \theta_j > y_0^j, \\ 0 & : \theta_j \leq y_0^j. \end{cases} \quad (27)$$

Thus, Eq. (21), the probability of a correct decision on trial $j - 1$, is needed to determine the mean decision time, DT_j , on trial j . The resulting values for c_j and DT_j can be used in Eq. (23) to determine the RR which again achieves a maximum for some choice of thresholds, $\theta_{1:n} := (\theta_1, \theta_2, \dots, \theta_n)$.

In Fig. 5 we show that the sequence of decision thresholds that maximize RR, $\theta_{1:n}^{\max}$, is decreasing ($\theta_1^{\max} \geq \theta_2^{\max} \geq \dots \geq \theta_n^{\max}$) across trials, consistent with our observation for two trials. Again, the increased accuracy in earlier trials improves accuracy in later trials, so there is value in gathering more information early in the trial sequence. Alternatively, an observer has less to gain from waiting for more information in a late trial, as the additional evidence can only be used on a few more trials. For intermediate to large values of ϵ , the observer uses a near constant threshold on the first $n - 1$ trials, and then a lower threshold on the last trial. As ϵ is decreased, the last threshold, θ_n^{\max} , collides with the boundary y_0^n , so the last decision is made instantaneously. The previous thresholds, $\theta_{n-1}^{\max}, \theta_{n-2}^{\max}, \dots$ collide with the corresponding boundaries $y_0^{n-1}, y_0^{n-2}, \dots$ in reverse order as ϵ is decreased further. Thus for lower values of ϵ , an increasing number of decisions toward the end of the sequence are made instantaneously.

6. Comparison to experimental results on repetition bias

We next discuss the experimental predictions of our model, motivated by previous observations of repetition bias. It has been long known that trends appear in the response sequences of subjects in series of TAFC trials (Fernberger, 1920), even when the trials are uncorrelated (Cho et al., 2002; Gao et al., 2009). For instance, when the state is the same on two adjacent trials, the probability of a correct response on the second trial increases (Fründ et al., 2014), and the response time decreases (Jones et al., 2013). In our model, if the assumed change rate of the environment does not match the true change rate, $\epsilon^{\text{assumed}} \neq \epsilon^{\text{true}}$, performance is decreased compared to $\epsilon^{\text{assumed}} = \epsilon^{\text{true}}$. However, this decrease may not be severe. Thus the observed response trends may arise partly due to misestimation of the transition rate ϵ , and the assumption that trials are dependent, even if they are not (Cho et al., 2002; Yu and Cohen, 2008).

Motivated by our findings on $RR_{1:n}^{\max}$ in the previous sections, we focus on the model of an observer who fixes their decision threshold across all trials, $\theta_j = \theta, \forall j$. Since allowing for dynamic thresholds does not impact $RR_{1:n}^{\max}$ considerably, we think the qualitative trends we observe here should generalize.

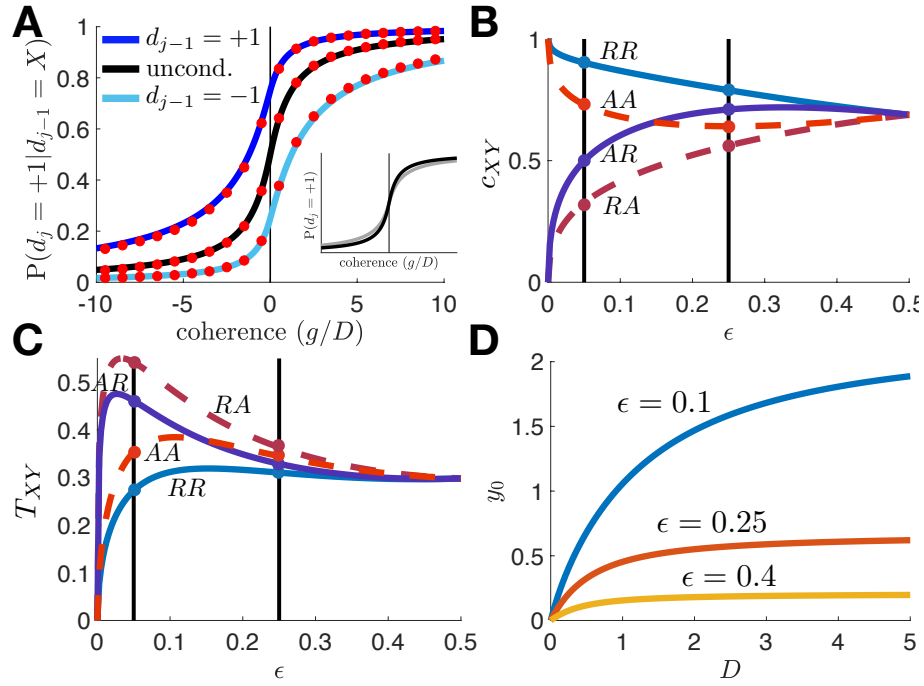


Figure 6: Repetition bias trends in the adaptive DDM. **A**. The psychometric functions in trial j conditioned on d_{j-1} shows that when $d_{j-1} = +1$, the fraction of $d_j = +1$ responses increases (dark blue). When $d_{j-1} = -1$, the fraction of $d_j = +1$ responses decreases (light blue). The weighted average of these two curves gives the unconditioned psychometric function (black curve). Coherence g/D is defined for drift $g \in \pm 1$, noise diffusion coefficient D , and assumed change rate $\epsilon = 0.25$. Inset: When $\epsilon^{\text{true}} = 0.5$ is the true change rate, the psychometric function is shallower (grey curve). **B**. Correct probability c_{XY} in current trial, plotted as a function of the assumed change rate ϵ , conditioned on the pairwise relations of the last three trials (R: repetitions; A: alternations). Note the ordering of c_{AA} and c_{AR} is exchanged as ϵ is increased, since the effects of the state two trials back are weakened. Compare colored dots at $\epsilon = 0.05$ and $\epsilon = 0.25$. **C**. Decision time T_{XY} conditioned on pairwise relations of last three trials. **D**. The initial condition, y_0 , increases as a function of the noise diffusion coefficient D for $\epsilon = 0.1, 0.25, 0.4$. In all cases, we set $T_D = 2$, $\theta = \theta_{\infty}^{\text{max}}$ for all trials, and $D = 1$ if not mentioned.

6.1. History-dependent psychometric functions

To provide a normative explanation of repetition bias, we compute the degree to which the optimal observer's choice on trial $j - 1$ biases their choice on trial j . For comparison with experimental work, we present the probability that $d_j = +1$, conditioned both on the choice on the previous trial, and on coherence³ (See Fig. 6A). In this case, $P(d_j = +1|g^j/D, d_{j-1} = \pm 1) = \pi_{\theta}(\pm y_0^j)$, where the exit probability $\pi_{\theta}(\cdot)$ is defined by Eq. (2), and the initial decision variable, y_0^j , is given by Eq. (14) using the threshold θ and the assumed transition rate, $\epsilon = \epsilon^{\text{assumed}}$. The unconditioned psychometric function is obtained by marginalizing over the true environmental change rate, ϵ^{true} , $P(d_j = +1|g^j/D) = (1 - \epsilon^{\text{true}})P(d_j = +1|g^j/D, d_{j-1} = +1) + \epsilon^{\text{true}}P(d_j = +1|g^j/D, d_{j-1} = -1)$, which equals $\pi_{\theta}(0)$ if $\epsilon = \epsilon^{\text{true}}$.

Fig. 6A shows that the probability of decision $d_j = +1$ is increased (decreased) if the same (opposite) decision was made on the previous trial. When $\epsilon^{\text{assumed}} \neq \epsilon^{\text{true}}$, the psychometric function is shallower and overall performance worsens (inset), although only slightly. Note that we have set $\theta = \theta_{\infty}^{\text{max}}$, so the decision threshold maximizes RR_{∞} in Eq. (25). The increased probability of a repetition arises from the fact that the initial condition $y^j(0)$ is biased in the direction of the previous decision.

6.2. Error rates and decision times

The adaptive DDM also exhibits repetition bias in decision times (Goldfarb et al., 2012). If the state at the current trial is the same as that on the previous trial (repetition) the decision time is shorter than average, while the decision

³We define coherence as the drift, $g^j \in \pm 1$, in trial j divided by the noise diffusion coefficient D (Gold and Shadlen, 2002; Bogacz et al., 2006).

time is increased if the state is different (alternation, see Appendix D). This occurs for optimal ($\epsilon^{\text{assumed}} = \epsilon^{\text{true}}$) and suboptimal ($\epsilon^{\text{assumed}} \neq \epsilon^{\text{true}}$) models as long as $\epsilon \in [0, 0.5)$. Furthermore, the rate of correct responses increases for repetitions and decrease for alternations (See Appendix D). However, the impact of the state on two or more previous trials on the choice in the present trial becomes more complicated. As in previous studies (Cho et al., 2002; Goldfarb et al., 2012; Jones et al., 2013), we propose that such repetition biases in the case of uncorrelated trials could be the result of the subject assuming $\epsilon = \epsilon^{\text{assumed}} \neq 0.5$, even when $\epsilon^{\text{true}} = 0.5$.

Following Goldfarb et al. (2012), we denote that three adjacent trial states as composed of repetitions (R), and alternations (A) (Cho et al., 2002; Goldfarb et al., 2012). Repetition-repetition (RR)⁴ corresponds to the three identical states in a sequence, $H^j = H^{j-1} = H^{j-2}$; RA corresponds to $H^j \neq H^{j-1} = H^{j-2}$; AR to $H^j = H^{j-1} \neq H^{j-2}$; and AA to $H^j \neq H^{j-1} \neq H^{j-2}$. The correct probabilities c_{XY} associated with all four of these cases can be computed explicitly, as shown in Appendix D. Here we assume that $\epsilon^{\text{assumed}} = \epsilon^{\text{true}}$ for simplicity, although it is likely that subject often use $\epsilon^{\text{assumed}} \neq \epsilon^{\text{true}}$ as shown in Cho et al. (2002); Goldfarb et al. (2012).

The probability of a correct response as a function of ϵ is shown for all four conditions in Fig. 6B. Clearly, c_{RR} is always largest, since the ideal observer correctly assumes repetitions are more likely than alternations (when $\epsilon \in [0, 0.5)$).

Note when $\epsilon = 0.05$, $c_{RR} > c_{AA} > c_{AR} > c_{RA}$ (colored dots), so a sequence of two alternations yields a higher correct probability on the current trial than a repetition preceded by an alternation. This may seem counterintuitive, since we might expect repetitions to always yield higher correct probabilities. However, two alternations yield the same state on the first and last trial in the three trial sequence ($H^j = H^{j-2}$). For small ϵ , since initial conditions begin close to threshold, then it is likely $d_j = d_{j-1} = d_{j-2}$, so if d_{j-2} is correct, then d_j will be too. At small ϵ , and at parameters we chose here, the increased likelihood of the first and last response being correct, outweighs the cost of the middle response likely being incorrect. Note that AA occurs with probability ϵ^2 , and is thus rarely observed when ϵ is small. As ϵ is increased, the ordering switches to $c_{RR} > c_{AR} > c_{AA} > c_{RA}$ (e.g., at $\epsilon = 0.25$), so a repetition in trial j leads to higher probability of a correct response. In this case, initial conditions are less biased, so the effects of the state two trials back are weaker. As before, we have fixed $\theta = \theta_{\infty}^{\text{max}}$.

These trends extend to decision times. Following a similar calculation as used for correct response probabilities, c_{XY} , we can compute the mean decision times, T_{XY} , conditioned on all possible three state sequences (See Appendix D). Fig. 6C shows that, as a function of ϵ , decision times following two repetitions, T_{RR} , are always the smallest. This is because the initial condition in trial j is most likely to start closest to the correct threshold: Decisions do not take as long when the decision variable starts close to the boundary. Also, note that when $\epsilon = 0.05$, $T_{RR} < T_{AA} < T_{AR} < T_{RA}$, and the explanation is similar to that for the probabilities of correct responses. When the environment changes slowly, decisions after two alternations will be quicker because a bias in the direction of the correct decision is more likely. The ordering is different at intermediate ϵ where $T_{RR} < T_{AR} < T_{AA} < T_{RA}$.

Notably, both orderings of the correct probability and decision times, as they depend on the three trial history, have been observed in previous studies of serial bias (Cho et al., 2002; Gao et al., 2009; Goldfarb et al., 2012): Sometimes $c_{AA} > c_{AR}$ and $T_{AA} < T_{AR}$ whereas sometimes $c_{AA} < c_{AR}$ and $T_{AA} > T_{AR}$. However, it appears to be more common that AA sequences lead to lower decision times, $T_{AA} < T_{AR}$ (e.g., See Fig. 3 in Goldfarb et al. (2012)). This suggests that subjects assume state switches are rare, corresponding to a smaller $\epsilon^{\text{assumed}}$.

6.3. Initial condition dependence on noise

Lastly, we explored how the initial condition, y_0 , which captures the information gathered on the previous trial, changes with signal coherence, g/D . We only consider the case $\theta = \theta_{\infty}^{\text{max}}$, and note that the threshold will vary with D . We find that the initial bias, y_0 , always tends to increase with D (Fig. 6D). The intuition for this result is that the optimal $\theta = \theta_{\infty}^{\text{max}}$ will tend to increase with D , since uncertainty increases with D , necessitating a higher threshold to maintain constant accuracy (See Eq. (4)). As the optimal threshold increases with D , the initial bias increases with it. This is consistent with experimental observations that show that bias tends to decrease with coherence, *i.e.* increase with noise amplitude (Fründ et al., 2014; Olianeghad et al., 2016; Braun et al., 2018).

We thus found that experimentally observed history-dependent perceptual biases can be explained using a model of an ideal observer accumulating evidence across correlated TAFC trials. Biases are stronger when the observer

⁴Note that RR here refers to repetition-repetition, as opposed to earlier when we used RR in Roman font to denote reward rate.

assumes a more slowly changing environment – smaller ϵ – where trials can have an impact on the belief of the ideal observer far into the future. Several previous studies have proposed this idea (Cho et al., 2002; Goldfarb et al., 2012; Braun et al., 2018), but to our knowledge, none have used a normative model to explain and quantify these effects.

7. Discussion

It has been known for nearly a century that observers take into account previous choices when making decisions (Fernberger, 1920). While a number of descriptive models have been used to explain experimental data (Fründ et al., 2014; Goldfarb et al., 2012), there have been no normative models that quantify how information accumulated on one trial impacts future decisions. Here we have shown that a straightforward, tractable extension of previous evidence-accumulation models can be used to do so.

To account for information obtained on one trial observers could adjust accumulation speed (Ratcliff, 1985; Diederich and Busemeyer, 2006; Urai et al., 2018), threshold (Bogacz et al., 2006; Goldfarb et al., 2012; Diederich and Busemeyer, 2006), or their initial belief on the subsequent trial (Bogacz et al., 2006; Braun et al., 2018). We have shown that an ideal observer adjusts their initial belief and decision threshold, but not the accumulation speed in a sequence of dependent, statistically-identical trials. Adjusting the initial belief increases reward rate by decreasing response time, but not the probability of a correct response. Changes in initial bias have a larger effect than adjustments in threshold, and there is evidence that human subjects do adjust their initial beliefs across trials (Fründ et al., 2014; Olianeghad et al., 2016; Braun et al., 2018). On the other hand, the effect of a dynamic threshold across trials diminishes as the number of trials increases: $RR_{1:n}^{\max}$ is nearly the same for constant thresholds θ as for when allowing dynamic thresholds $\theta_{1:n}$. Given the cognitive cost of trial-to-trial threshold adjustments, it is thus unclear whether human subjects would implement such a strategy (Balci et al., 2011).

To isolate the effect of previously gathered information, we have assumed that the observers do not receive feedback or reward until the end of the trial. Observers can thus use only information gathered on previous trials to adjust their bias on the next. However, our results can be easily extended to the case when feedback is provided on each trial. The expression for the initial condition y_0 is simpler in this case, since feedback provides certainty concerning the state on the previous trial.

We also assumed that an observer knows the length of the trial sequence, and uses this number to optimize reward rate. It is rare that an organism would have such information in a natural situation. However, our model can be extended to random sequence durations if we allow observers to marginalize over possible sequence lengths. If decision thresholds are fixed to be constant across all trials $\theta_j = \theta$, the projected RR is the average number of correct decisions $c\langle n \rangle$ divided by the average time of the sequence. For instance, if the trial length follows a geometric distribution, $n \sim \text{Geom}(p)$, then

$$\bar{RR}(\theta; \epsilon) = \frac{1/p}{[1 - e^{-\theta/D}][\theta/p - (1/p - 1)(1 - 2\epsilon)y_0] + (T_D/p)[1 + e^{-\theta/D}]}$$

This equation has the same form as Eq. (24), the RR in the case of fixed trial number n , with n replaced by the average $\langle n \rangle = 1/p$. Thus, we expect that as p decreases, and the average trial number, $1/p$, increases, resulting in an increase in the RR and larger optimal decision thresholds θ^{\max} (as in Fig. 4). The case of dynamic thresholds is somewhat more involved, but can be treated similarly. However, as the number of trials is generated by a memoryless process, observers gain no further information about how many trials remain after the present, and will thus not adjust their thresholds as in Section 5.3.

Our conclusions also depend on the assumptions the observer makes about the environment. For instance, an observer may not know the exact probability of change between trials, so that $\epsilon^{\text{assumed}} \neq \epsilon^{\text{true}}$. Indeed, there is evidence that human subjects adjust their estimate of ϵ across a sequence of trials, and assume trials are correlated, even when they are not (Cho et al., 2002; Gao et al., 2009; Yu and Cohen, 2008). Our model can be expanded to a hierarchical model in which both the states $H^{1:n}$ and the change rate, ϵ , are inferred across a sequence of trials (Radillo et al., 2017). This would result in model ideal observers that use the current estimate of the ϵ to set the initial bias on a trial (Yu and Cohen, 2008). A similar approach could also allow us to model observers that incorrectly learn, or make wrong assumptions, about the rate of change.

Modeling approaches can also point to the neural computations that underlie the decision-making process. Cortical representations of previous trials have been identified (Nogueira et al., 2017; Akrami et al., 2018), but it is unclear how

the brain learns and combines latent properties of the environment (the change rate ϵ), with sensory information to make inferences. Recent work on repetition bias in working memory suggests short-term plasticity serves to encode priors in neural circuits, which bias neural activity patterns during memory retention periods (Papadimitriou et al., 2015; Kilpatrick, 2018). Such synaptic mechanisms may also allow for the encoding of previously accumulated information in sequential decision-making tasks, even in cases where state sequences are not correlated. Indeed, evolution may imprint characteristics of natural environments onto neural circuits, shaping whether and how ϵ is learned in experiments and influencing trends in human response statistics (Fawcett et al., 2014; Glaze et al., 2018). Ecologically adapted decision rules may be difficult to train away, especially if they do not impact performance too adversely (Todd and Gigerenzer, 2007).

The normative model we have derived can be used to identify task parameter ranges that will help tease apart the assumptions made by subjects (Cho et al., 2002; Gao et al., 2009; Goldfarb et al., 2012). Since our model describes the behavior of an ideal observer, it can be used to determine whether experimental subjects use near-optimal or sub-optimal evidence-accumulation strategies. Furthermore, our adaptive DDM opens a window to further instantiations of the DDM that consider the influence of more complex state-histories on the evidence-accumulation strategies adopted within a trial. Uncovering common assumptions about state-histories will help guide future experiments, and help us better quantify the biases and core mechanisms of human decision-making.

Acknowledgements

This work was supported by an NSF/NIH CRCNS grant (R01MH115557). KN was supported by an NSF Graduate Research Fellowship. ZPK was also supported by an NSF grant (DMS-1615737). KJ was also supported by NSF grant DBI-1707400 and DMS-1517629.

Appendix A. Derivation of the drift-diffusion model for a single trial

We assume that an optimal observer integrates a stream of noisy measurements $\xi_{1:s}^1 = (\xi_1^1, \xi_2^1, \dots, \xi_s^1)$ at equally spaced times $t_{1:s} = (t_1, t_2, \dots, t_s)$ (Wald and Wolfowitz, 1948; Bogacz et al., 2006). The likelihood functions, $f_{\pm}(\xi) := P(\xi|H_{\pm})$, define the probability of each measurement, ξ , conditioned on the environmental state, H_{\pm} . Observations in each trial are combined with the prior probabilities $P(H_{\pm})$. We assume a symmetric prior, $P(H_{\pm}) = 1/2$. The probability ratio is then

$$R_s^1 = \frac{P(H^1 = H_+ | \xi_{1:s}^1)}{P(H^1 = H_- | \xi_{1:s}^1)} = \frac{f_+(\xi_1^1)f_+(\xi_2^1) \cdots f_+(\xi_s^1)P(H^1 = H_+)}{f_-(\xi_1^1)f_-(\xi_2^1) \cdots f_-(\xi_s^1)P(H^1 = H_-)}, \quad (\text{A.1})$$

due to the independence of each measurement ξ_s^1 . Thus, if $R_s^1 > 1$ ($R_s^1 < 1$) then $H^1 = H_+$ ($H^1 = H_-$) is the more likely state. Eq. (A.1) can be written recursively

$$R_s^1 = \left(\frac{f_+(\xi_s^1)}{f_-(\xi_s^1)} \right) R_{s-1}^1, \quad (\text{A.2})$$

where $R_0^1 = P(H_+)/P(H_-) = 1$, due to the assumed prior at the beginning of trial 1.

Taking the logarithm of Eq. (A.2) allows us to express the recursive relation for the log-likelihood ratio (LLR) $L_s^1 = \ln R_s^1$ as an iterative sum

$$L_s^1 = L_{s-1}^1 + \ln \frac{f_+(\xi_s^1)}{f_-(\xi_s^1)}, \quad (\text{A.3})$$

where $L_0^1 = \ln [P(H_+)/P(H_-)]$, so if $L_s^1 \geq 0$ then $H^1 = H_+$ is the more likely state. Taking the continuum limit $\Delta t \rightarrow 0$ of the timestep $\Delta t := t_s - t_{s-1}$, we can use the functional central limit theorem to yield the DDM (See Bogacz et al. (2006); Veliz-Cuba et al. (2016) for details):

$$dy^1 = g^1 dt + \sqrt{2D}dW, \quad (\text{A.4})$$

where W is a Wiener process, the drift $g^1 \in g_{\pm} = \frac{1}{\Delta t} E_{\xi} \left[\ln \frac{f_{+}(\xi)}{f_{-}(\xi)} | H_{\pm} \right]$ depends on the state, and $2D = \frac{1}{\Delta t} \text{Var}_{\xi} \left[\ln \frac{f_{+}(\xi)}{f_{-}(\xi)} | H_{\pm} \right]$ is the variance which depends only on the noisiness of each observation, but not the state.

With Eq. (1) in hand, we can relate $y^1(t)$ to the probability of either state H_{\pm} by noting its Gaussian statistics in the case of free boundaries

$$P(y^1(t) | H^1 = H_{\pm}) = \frac{1}{\sqrt{4\pi Dt}} e^{-(y^1(t) \mp t)^2 / (4Dt)},$$

so that

$$\frac{P(H^1 = H_{+} | y^1(t))}{P(H^1 = H_{-} | y^1(t))} = \frac{P(y^1(t) | H^1 = H_{+})}{P(y^1(t) | H^1 = H_{-})} = e^{y^1(t)/D}. \quad (\text{A.5})$$

Note, an identical relation is obtained in Bogacz et al. (2006) in the case of absorbing boundaries. Either way, it is clear that $y^1 = D \cdot \text{LLR}_1$. This means that before any observations have been made $e^{y^1(0)/D} = \frac{P(H^1=H_{+})}{P(H^1=H_{-})}$, so we scale the log ratio of the prior by D to yield the initial condition y_0^1 in Eq. (1).

Appendix B. Threshold determines probability of being correct

We note that Eq. (A.5) extends to any trial j in the case of an ideal observer. In this case, when the decision variable $y^2(T_2) = \theta$, we can write

$$\frac{P(H^2 = H_{+} | y^2(T_2) = \theta)}{P(H^2 = H_{-} | y^2(T_2) = \theta)} = e^{\theta/D}, \quad (\text{B.1})$$

so that a rearrangement of Eq. (B.1) yields

$$c_2 := \frac{P(H^2 = H_{+} | y^2(T_2) = \theta) P(y^2(T_2) = \theta)}{P(d_2 = +1)} = \frac{1}{1 + e^{-\theta/D}} = c_1, \quad (\text{B.2})$$

Therefore the probability of correct decisions on trials 1 and 2 are equal, and determined by $\theta = \theta_{1,2}$, and D . A similar computation shows that the threshold, $\theta_j = \theta$, and diffusion rate, D , *but not the initial belief*, determine the probability of a correct response on any trial j .

To demonstrate self-consistency in the general case of decision thresholds that may differ between trials $\theta_{1,2}$, we can also derive the formula for c_2 in the case $y_0^2 \leq \theta_2$. Our alternative derivation of Eq. (B.2) begins by computing the total probability of a correct choice on trial 2 by combining the conditional probabilities of a correct choice given decision $d^1 = +1$ leading to initial belief y_0^2 , and decision $d^1 = -1$ leading to initial belief $-y_0^2$:

$$\begin{aligned} c_2 &= [(1 - \epsilon)c_1 + \epsilon(1 - c_1)] \cdot \pi_{\theta_2}(y_0^2) + [\epsilon c_1 + (1 - \epsilon)(1 - c_1)] \cdot \pi_{\theta_2}(-y_0^2), \\ &= \frac{(1 - \epsilon)e^{\theta_1/D} + \epsilon(1 - \epsilon)(e^{\theta_1/D} - e^{-\theta_2/D}) + \epsilon(1 - e^{(\theta_1 - \theta_2)/D})}{e^{\theta_1/D} + 1} \frac{1}{((1 - \epsilon)e^{\theta_1/D} + \epsilon)(1 - e^{-2\theta_2/D})} \\ &\quad + \frac{\epsilon e^{\theta_1/D} + (1 - \epsilon)\epsilon(e^{\theta_1/D} - e^{-\theta_2/D}) + (1 - \epsilon)(1 - e^{(\theta_1 - \theta_2)/D})}{e^{\theta_1/D} + 1} \frac{1}{(\epsilon e^{\theta_1/D} + (1 - \epsilon))(1 - e^{-2\theta_2/D})} \\ &= \frac{(\epsilon e^{\theta_1/D} + (1 - \epsilon))((1 - \epsilon)e^{\theta_1/D} + \epsilon)(e^{\theta_1/D} + 1)(1 - e^{-\theta_2/D})}{(\epsilon e^{\theta_1/D} + (1 - \epsilon))((1 - \epsilon)e^{\theta_1/D} + \epsilon)(e^{\theta_1/D} + 1)(1 - e^{-2\theta_2/D})} \\ c_2 &= \frac{1}{1 + e^{-\theta_2/D}}, \end{aligned} \quad (\text{B.3})$$

showing again that the correct probability c_2 on trial 2 is solely determined by θ_2 and D , assuming $y_0^2 \leq \theta_2$.

Appendix C. Mean response time on the second trial

We can compute the average time until a decision on trial 2 by marginalizing over $d_1 = \pm 1$,

$$DT_2|(H^2 = H_{\pm}) = \sum_{s=\pm 1} T(s \cdot y_0^2; \theta_2) \cdot P(d_1 = s|H^2 = H_{\pm}), \quad (\text{C.1})$$

since $d_1 = \pm 1$ will lead to the initial condition $\pm y_0^2$ in trial 2. We can compute $T(\pm y_0^2; \theta)$ using Eq. (7) as

$$T(\pm y_0^2; \theta_2) = \theta_2 \left[\coth \frac{\theta_2}{D} - \operatorname{csch} \frac{\theta_2}{D} \left[\frac{\epsilon e^{\theta_1/D} + (1 - \epsilon)}{(1 - \epsilon) e^{\theta_1/D} + \epsilon} \right]^{\pm 1} \right] \mp D \ln \frac{(1 - \epsilon) e^{\theta_1/D} + \epsilon}{\epsilon e^{\theta_1/D} + (1 - \epsilon)}. \quad (\text{C.2})$$

The second terms in the summands of Eq. (C.1) are given by

$$P(d_1 = \pm 1|H^2 = H_{\pm}) = \frac{P(H^2 = H_{\pm}|d_1 = \pm 1)P(d_1 = \pm 1)}{P(H^2 = H_{\pm})} = (1 - \epsilon)c_1 + \epsilon(1 - c_1),$$

$$P(d_1 = \mp 1|H^2 = H_{\pm}) = \frac{P(H^2 = H_{\pm}|d_1 = \mp 1)P(d_1 = \mp 1)}{P(H^2 = H_{\pm})} = \epsilon c_1 + (1 - \epsilon)(1 - c_1),$$

and note $DT_2 = [DT_2|H^2 = H_+]P(H^2 = H_+) + [DT_2|H^2 = H_-]P(H^2 = H_-) = DT_2|(H^2 = H_{\pm})$.

Using Eqs. (C.1-C.2) and simplifying assuming $\theta_{1,2} = \theta$, we obtain Eq. (17). For unequal thresholds, $\theta_1 \neq \theta_2$, a similar computation yields Eq. (19).

Appendix D. Repetition-dependence of correct probabilities and decision times

Here we derive formulas for correct probabilities c_j and average decision times DT_j in trial j , conditioned on whether the current trial is a repetition ($R: H^j = H^{j-1}$) or alternation ($A: H^j \neq H^{j-1}$) compared with the previous trial. To begin, we first determine the correct probability, conditioned on repetitions (R), $c_j|R := c_j|(H^j = H^{j-1} = H_{\pm})$:

$$\begin{aligned} c_j|R &= \sum_{s=\pm 1} \pi_{\pm\theta}(\pm s \cdot y_0^j) \cdot P(d_{j-1} = \pm s|H^j = H^{j-1} = H_{\pm}) \\ &= \frac{1 - e^{-\theta/D} e^{-y_0^j/D}}{1 - e^{-2\theta/D}} \cdot \frac{1}{1 + e^{-\theta/D}} + \frac{1 - e^{-\theta/D} e^{+y_0^j/D}}{1 - e^{-2\theta/D}} \cdot \frac{e^{-\theta/D}}{1 + e^{-\theta/D}} \\ &= \frac{(1 - \epsilon)e^{\theta/D}}{[1 + e^{-\theta/D}][(1 - \epsilon)e^{\theta/D} + \epsilon]} + \frac{\epsilon}{[1 + e^{-\theta/D}][\epsilon e^{\theta/D} + (1 - \epsilon)]} \\ &= \frac{(1 - \epsilon)\epsilon e^{2\theta/D} + (1 - \epsilon)e^{\theta/D} + \epsilon^2}{[1 + e^{-\theta/D}][(1 - \epsilon)e^{\theta/D} + \epsilon][\epsilon e^{\theta/D} + (1 - \epsilon)]}. \end{aligned}$$

In the case of alternations, $c_j|A := c_j|(H^j \neq H^{j-1}) = c_j|(H^j \neq H^{j-1} = H_{\pm})$:

$$\begin{aligned} c_j|A &= \sum_{s=\pm 1} \pi_{\mp\theta}(\mp s \cdot y_0^j) \cdot P(d_{j-1} = \mp s|H^j \neq H^{j-1} = H_{\pm}) \\ &= \frac{1 - e^{-\theta/D} e^{-y_0^j/D}}{1 - e^{-2\theta/D}} \cdot \frac{e^{-\theta/D}}{1 + e^{-\theta/D}} + \frac{1 - e^{-\theta/D} e^{+y_0^j/D}}{1 - e^{-2\theta/D}} \cdot \frac{1}{1 + e^{-\theta/D}} \\ &= \frac{(1 - \epsilon)}{[1 + e^{-\theta/D}][(1 - \epsilon)e^{\theta/D} + \epsilon]} + \frac{\epsilon e^{\theta/D}}{[1 + e^{-\theta/D}][\epsilon e^{\theta/D} + (1 - \epsilon)]} \\ &= \frac{(1 - \epsilon)\epsilon e^{2\theta/D} + \epsilon e^{\theta/D} + (1 - \epsilon)^2}{[1 + e^{-\theta/D}][(1 - \epsilon)e^{\theta/D} + \epsilon][\epsilon e^{\theta/D} + (1 - \epsilon)]}. \end{aligned}$$

Thus, we can show that the correct probability in the case of repetitions is higher than that for alternations, $c_j|R > c_j|A$ for any $\theta > 0$ and $\epsilon \in [0, 0.5)$, since

$$c_j|R - c_j|A = \frac{(1 - 2\epsilon)[e^{\theta/D} - 1]}{[1 + e^{-\theta/D}][(1 - \epsilon)e^{\theta/D} + \epsilon][\epsilon e^{\theta/D} + (1 - \epsilon)]} > 0.$$

Now, we also demonstrate that decision times are shorter for repetitions than for alternations. First, note that when the current trial is a repetition, we can again marginalize to compute $DT_j|R := DT_j|(H^j = H^{j-1} = H_{\pm})$:

$$DT_j|R = T(+y_0^j; \theta) \frac{1}{1 + e^{-\theta/D}} + T(-y_0^j; \theta) \frac{e^{-\theta/D}}{1 + e^{-\theta/D}},$$

whereas in the case of alternations

$$DT_j|A = T(+y_0^j; \theta) \frac{e^{-\theta/D}}{1 + e^{-\theta/D}} + T(-y_0^j; \theta) \frac{1}{1 + e^{-\theta/D}}.$$

By subtracting $DT_j|A$ from $DT_j|R$, we can show that $DT_j|R < DT_j|A$ for all $\theta > 0$ and $\epsilon \in [0, 0.5)$. To begin, note that

$$DT_j|R - DT_j|A = \frac{1 - e^{-\theta/D}}{1 + e^{-\theta/D}} [T(+y_0^j; \theta) - T(-y_0^j; \theta)].$$

Since it is clear the first term in the product above is positive for $\theta, D > 0$, we can show $DT_j|R - DT_j|A < 0$ and thus $DT_j|R < DT_j|A$ if $T(+y_0^j; \theta) < T(-y_0^j; \theta)$. We verify this is so by first noting the function $F(y) = (e^{y/D} - e^{-y/D})/y$ is always nondecreasing since $F'(y) = \left[\frac{y}{D} \cosh \frac{y}{D} - \sinh \frac{y}{D} \right] / y^2 \geq 0$ since $y/D \geq \tanh y/D$. In fact $F(y)$ is strictly increasing everywhere but when $y = 0$. This means that as long as $y_0 < \theta$, $F(y_0) < F(\theta)$, which implies

$$T(+y_0^j; \theta) - T(-y_0^j; \theta) = 2\theta \frac{e^{y_0/D} - e^{-y_0/D}}{e^{\theta/D} - e^{-\theta/D}} - 2y_0 < 0,$$

so $T(+y_0^j; \theta) < T(-y_0^j; \theta)$ as we wished to show, so $DT_j|R < DT_j|A$.

Moving to three state sequences, we note that we can use our formulas for $c_j|R$ and $c_j|A$ in Appendix D as well as $\pi_{\theta}(y)$ described by Eq. (2) to compute

$$\begin{aligned} c_{RR} &:= c_j|RR = \pi_{\pm\theta}(\pm y_0^j) \cdot c_j|R + \pi_{\pm\theta}(\mp y_0^j) \cdot (1 - c_j)|R, \\ c_{RA} &:= c_j|RA = \pi_{\pm\theta}(\mp y_0^j) \cdot c_j|R + \pi_{\pm\theta}(\pm y_0^j) \cdot (1 - c_j)|R, \\ c_{AR} &:= c_j|AR = \pi_{\pm\theta}(\pm y_0^j) \cdot c_j|A + \pi_{\pm\theta}(\mp y_0^j) \cdot (1 - c_j)|A, \\ c_{AA} &:= c_j|AA = \pi_{\pm\theta}(\mp y_0^j) \cdot c_j|A + \pi_{\pm\theta}(\pm y_0^j) \cdot (1 - c_j)|A. \end{aligned}$$

Applying similar logic to the calculation of the decision times as they depend on the three state sequences, we find that

$$\begin{aligned} T_{RR} &:= DT_j|RR = T(+y_0^j; \theta) \cdot c_j|R + T(-y_0^j; \theta) \cdot (1 - c_j)|R, \\ T_{RA} &:= DT_j|RA = T(-y_0^j; \theta) \cdot c_j|R + T(+y_0^j; \theta) \cdot (1 - c_j)|R, \\ T_{AR} &:= DT_j|AR = T(+y_0^j; \theta) \cdot c_j|A + T(-y_0^j; \theta) \cdot (1 - c_j)|A, \\ T_{AA} &:= DT_j|AA = T(-y_0^j; \theta) \cdot c_j|A + T(+y_0^j; \theta) \cdot (1 - c_j)|A. \end{aligned}$$

Appendix E. Numerical simulations

All numerical simulations of the DDM were performed using the Euler-Maruyama method with a timestep $dt = 0.005$, computing each point from 10^5 realizations. To optimize reward rates, we used the Nelder-Mead simplex method (`fminsearch` in MATLAB). All code used to generate figures will be available on [github](https://github.com).

Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M., Gardner, J. L., 2016. Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences* 113 (25), E3548–E3557.

Akrami, A., Kopec, C. D., Diamond, M. E., Brody, C. D., 2018. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature* 554 (7692), 368.

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., Cohen, J. D., 2011. Acquisition of decision making criteria: reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics* 73 (2), 640–657.

- Bertelson, P., 1961. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology* 13 (2), 90–102.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., Cohen, J. D., 2006. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review* 113 (4), 700.
- Bogacz, R., Hu, P. T., Holmes, P. J., Cohen, J. D., 2010. Do humans produce the speed–accuracy trade-off that maximizes reward rate? *Quarterly journal of experimental psychology* 63 (5), 863–891.
- Braun, A., Urai, A. E., Donner, T. H., Jan. 2018. Adaptive History Biases Result from Confidence-weighted Accumulation of Past Choices. *Journal of Neuroscience*.
- Brody, C. D., Hanks, T. D., 2016. Neural underpinnings of the evidence accumulator. *Current opinion in neurobiology* 37, 149–157.
- Brunton, B. W., Botvinick, M. M., Brody, C. D., 2013. Rats and humans can optimally accumulate evidence for decision-making. *Science* 340 (6128), 95–98.
- Chittka, L., Skorupski, P., Raine, N. E., 2009. Speed–accuracy tradeoffs in animal decision making. *Trends in ecology & evolution* 24 (7), 400–407.
- Cho, R. Y., Nystrom, L. E., Brown, E. T., Jones, A. D., Braver, T. S., Holmes, P. J., Cohen, J. D., 2002. Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective, & Behavioral Neuroscience* 2 (4), 283–299.
- Corcoran, A. J., Conner, W. E., 2016. How moths escape bats: predicting outcomes of predator–prey interactions. *Journal of Experimental Biology* 219 (17), 2704–2715.
- Dehaene, S., Sigman, M., 2012. From a single decision to a multi-step algorithm. *Current opinion in neurobiology* 22 (6), 937–945.
- Diederich, A., Busemeyer, J. R., 2006. Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics* 68 (2), 194–207.
- Fawcett, T. W., Fallenstein, B., Higginson, A. D., Houston, A. I., Mallpress, D. E., Trimmer, P. C., McNamara, J. M., et al., 2014. The evolution of decision rules in complex environments. *Trends in cognitive sciences* 18 (3), 153–161.
- Fernberger, S. W., 1920. Interdependence of judgments within the series for the method of constant stimuli. *Journal of Experimental Psychology* 3 (2), 126.
- Fründ, I., Wichmann, F. A., Macke, J. H., Jun. 2014. Quantifying the effect of intertrial dependence on perceptual decisions. *Journal of Vision* 14 (7).
- Gao, J., Wong-Lin, K., Holmes, P., Simen, P., Cohen, J. D., 2009. Sequential effects in two-choice reaction time tasks: decomposition and synthesis of mechanisms. *Neural Computation* 21 (9), 2407–2436.
- Gardiner, C., 2009. *Stochastic methods*. Vol. 4. Springer Berlin.
- Glaze, C. M., Filipowicz, A. L., Kable, J. W., Balasubramanian, V., Gold, J. I., 2018. A bias–variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour* 2 (3), 213.
- Gold, J. I., Shadlen, M. N., 2002. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* 36 (2), 299–308.
- Gold, J. I., Shadlen, M. N., 2007. The neural basis of decision making. *Annual review of neuroscience* 30.
- Goldfarb, S., Wong-Lin, K., Schwemmer, M., Leonard, N. E., Holmes, P., 2012. Can post-error dynamics explain sequential reaction time patterns? *Frontiers in Psychology* 3, 213.
- Gordon, D. M., 1991. Behavioral flexibility and the foraging ecology of seed-eating ants. *The American Naturalist* 138 (2), 379–411.
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C. F., 2005. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310 (5754), 1680–1683.
- Jones, M., Curran, T., Mozer, M. C., Wilder, M. H., 2013. Sequential effects in response time reveal learning mechanisms and event representations. *Psychological review* 120 (3), 628.
- Kilpatrick, Z. P., 2018. Synaptic mechanisms of interference in working memory. *Scientific Reports* 8 (1), 7879.
URL <https://doi.org/10.1038/s41598-018-25958-9>
- Kirby, N. H., 1976. Sequential effects in two-choice reaction time: Automatic facilitation or subjective expectancy? *Journal of Experimental Psychology: Human perception and performance* 2 (4), 567.
- Latimer, K. W., Yates, J. L., Meister, M. L., Huk, A. C., Pillow, J. W., 2015. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science* 349 (6244), 184–187.
- Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., Preis, T., 2013. Quantifying wikipedia usage patterns before stock market moves. *Scientific reports* 3, 1801.
- Nogueira, R., Abolafia, J. M., Drugowitsch, J., Balaguer-Ballester, E., Sanchez-Vives, M. V., Moreno-Bote, R., 2017. Lateral orbitofrontal cortex anticipates choices and integrates prior with current information. *Nature Communications* 8, 14823.
- Olianezhad, F., Tohidi-Moghaddam, M., Zabbah, S., Ebrahimpour, R., Nov. 2016. Residual Information of Previous Decision Affects Evidence Accumulation in Current Decision. [arXiv.org](https://arxiv.org/abs/1611.08881).
- Papadimitriou, C., Ferdoash, A., Snyder, L. H., 2015. Ghosts in the machine: memory interference from the previous trial. *Journal of neurophysiology* 113 (2), 567–577.
- Pashler, H., Baylis, G. C., 1991. Procedural learning: II. intertrial repetition effects in speeded-choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17 (1), 33.
- Radillo, A. E., Veliz-Cuba, A., Josić, K., Kilpatrick, Z. P., 2017. Evidence accumulation and change rate inference in dynamic environments. *Neural computation* 29 (6), 1561–1610.
- Raposo, D., Sheppard, J. P., Schrater, P. R., Churchland, A. K., 2012. Multisensory decision-making in rats and humans. *Journal of neuroscience* 32 (11), 3726–3735.
- Ratcliff, R., 1978. A theory of memory retrieval. *Psychological review* 85 (2), 59.
- Ratcliff, R., 1985. Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological review* 92 (2), 212.
- Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation* 20 (4), 873–922.
- Ratcliff, R., Smith, P. L., 2004. A comparison of sequential sampling models for two-choice reaction time. *Psychological review* 111 (2), 333.
- Schall, J. D., 2001. Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience* 2 (1), 33.

- Seeley, T. D., Camazine, S., Sneyd, J., 1991. Collective decision-making in honey bees: how colonies choose among nectar sources. *Behavioral Ecology and Sociobiology* 28 (4), 277–290.
- Shadlen, M. N., Newsome, W. T., 2001. Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of neurophysiology* 86 (4), 1916–1936.
- Stevenson, P. A., Rillich, J., 2012. The decision to fight or flee—insights into underlying mechanism in crickets. *Frontiers in neuroscience* 6, 118.
- Stone, M., 1960. Models for choice-reaction time. *Psychometrika* 25 (3), 251–260.
- Todd, P. M., Gigerenzer, G., 2007. Environments that make us smart: Ecological rationality. *Current directions in psychological science* 16 (3), 167–171.
- Urai, A. E., de Gee, J. W., Donner, T. H., 2018. Choice history biases subsequent evidence accumulation. *bioRxiv*, 251595.
- Veliz-Cuba, A., Kilpatrick, Z. P., Josić, K., 2016. Stochastic models of evidence accumulation in changing environments. *SIAM Review* 58 (2), 264–289.
- Wald, A., Wolfowitz, J., 1948. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 326–339.
- Yu, A. J., Cohen, J. D., 2008. Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems* 21, 1873–1880.