# Intertrial correlations in sequential decision-making tasks

Khanh Nguyen[1,*], Krešmir Josić[1], and Zachary P Kilpatrick[2]
1: University of Houston; 2: University of Colorado Boulder; *kpnguyen@math.uh.edu

Two alternative forced choice (TAFC) tasks have been used extensively to identify strategies humans and other animals use to accumulate evidence. Experiments demonstrated that subjects can learn the latent probabilistic structure of their environment[3,5,6], improving decision performance. For instance, humans can learn to account for dependencies between trials when the correct choice has a greater than chance probability of being repeated. We derive and analyze normative probabilistic models of evidence accumulation in such situations. Within each trial, the belief of an observer is described by a drift-diffusion process with an initial condition that depends on the trial history.

We build on the standard drift-diffusion model (DDM), which has been used to describe behavioral and neural data in TAFC tasks[1,4]. An observer's belief on the $n^{\text{th}}$ trial obeys a normalized DDM[1], $\mathrm{d}y^n = s_n \cdot \mathrm{d}t + \sigma \cdot \mathrm{d}W$, where $s_n \in \pm 1$ is the correct choice on the current trial ($H^n \in H_\pm$) and $\sigma$ is related to the uncertainty (noise) in individual observations. The observer makes its $n^{\text{th}}$ decision when $y^n$ crosses one of the two thresholds, $y^n(t) = \pm\theta_n$ (Fig. 1). It can be shown that the probability of a correct decision on trial $n$ is $c_n = 1/(1+e^{-2\theta_n})$. A large body of work shows that this optimal model describes evidence integration for a range of tasks and species [1,2,4].

A key feature of our work is that the correct choice on each trial $H^n$ switches with lower than chance probability ($\epsilon := \Pr(H^{n+1} \neq H^n) < 0.5$), according to a two-state Markov process. An optimal observer therefore must set their initial belief $y^{n+1}(0) = \pm y_0(\theta_n) \neq 0$, depending on the decision on trial $n$. This carries forward information about the most likely state ($H^n \in H_\pm$) on trial $n$. Using principles of probabilistic inference, we find that if the response on trial $n$ is $H_\pm$, then the observer should set $\pm y_0(\theta_n) = \pm \ln \dfrac{(1-\epsilon)\mathrm{e}^{\theta_n} + \epsilon}{\epsilon \mathrm{e}^{\theta_n} + (1-\epsilon)}$.

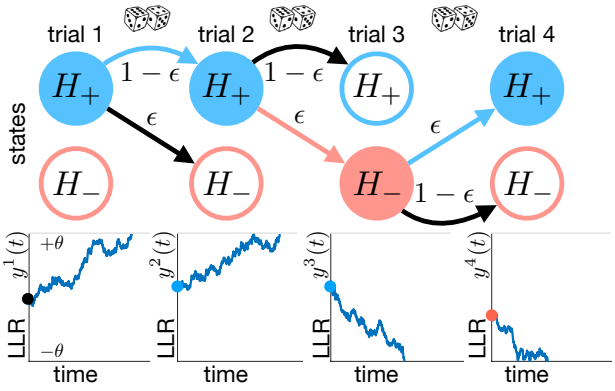


Figure 1: The correct choice on each trial $H^n \in H_\pm$ switches stochastically between trials with probability $\epsilon$. An observer accumulates noisy evidence within trial $n+1$, starting at $y^{n+1}(0) = \pm y_0(\theta_n)$ if $y^n(t_{\text{end}}) = \pm\theta_n$.

Predicted performance of the observer on trial $n$ is calculated via the reward rate (RR) function $\mathrm{RR}_n = \dfrac{c_n}{E(T_n) + T_I}$, where $c_n$ is the probability of a correct decision, $T_I$ is the intertrial interval, and the expected decision time is

$$E(T_n) = [c_{n-1}(1-\epsilon) + (1-c_{n-1})\epsilon]T^+(\theta_{n-1}, \theta_n) + [c_{n-1}\epsilon + (1-c_{n-1})(1-\epsilon)]T^-(\theta_{n-1}, \theta_n) \quad (1)$$

where $T^+$ ($T^-$) is the mean time to a decision if the current and anticipated states of the environment $H^n$ are the same (different). Mean first passage times are determined by analyzing the corresponding Fokker Planck equation[1]. Each contribution is weighted by the probability of the true and anticipated states matching or not. Note, $E(T_1)$ is the mean time for a DDM to cross $\pm\theta_1$ given $y^1(0) = 0$.

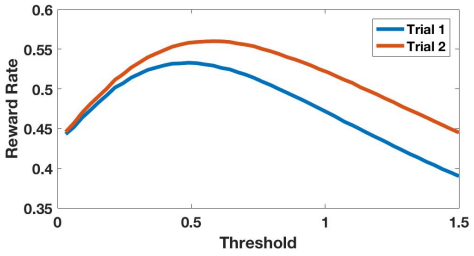


Figure 2: Reward Rate (RR) as a function of thresholds $\theta_{1,2}$ for the first and second trials, showing the RR increases in trial 2, due to information obtained from trial 1. Here $\epsilon = 0.25$.

To understand the impact of propagating information across trials, we can study the RR for trials 1 and 2 as a function of the threshold $\theta$ (Fig. 2). The RR always increases from trial 1 to 2 for $0 \leq \epsilon < 0.5$ and $\theta_1, \theta_2 > 0$. If we set $\theta_1 = \theta_2$, so that $c_1 = c_2$, the time to a decision is less for trial 2 than for trial 1, $E(T_2) < E(T_1)$. Note also that the RR peaks in trial 2 at a larger threshold value ($\theta_2^{\text{opt}} > \theta_1^{\text{opt}}$). Our initial analysis suggests that an optimal observer should increase their decision threshold from one trial to the next to maximize their long term reward rate ($\theta_{n+1}^{\text{opt}} > \theta_n^{\text{opt}}$). We have also found that sequential memory across trials impacts performance more strongly on protocols with smaller values of $\epsilon$. We conclude that optimal evidence accumulation for correlated sequential TAFC task trials should involve a history-dependent modification of initial conditions, $y^n(0)$, and graduated thresholds, $\theta_n$.

**References:** [1] R. Bogacz *et al.*, *Psychol. Rev.*, 113(4):700-765, 2006. [2] B. Brunton *et al.*, *Science*, 340(6128):95-98, 2013. [3] R. Cho *et al.*, *Cogn. Affect. Behave. Ne.*, 2(4):283-299, 2002. [4] J. Gold *et al.*, *Annu. Rev. Neurosci.*, 30:535-574, 2007 [5] S. Goldfarb, *et al.*, *Front. Psychol.*, 3, 2012. [6] T. Kim, *et al.*, *J. Neurosci.*, 37(13):3632-3645, 2017. [7] A. Radillo, *et al.*, *Neural Comput.* 29: 1561-1610 (2017).