

How Trial Correlations and Feedback Shape Sequential Decision-Making

by

Khanh P Nguyen

A Dissertation submitted to the Department of Mathematics,
College of Natural Sciences and Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Mathematics

Chair of Committee: Krešimir Josić

Committee Member: Zachary Kilpatrick

Committee Member: William Ott

Committee Member: Mikhail Perepelitsa

University of Houston
August 2020

Copyright 2020, Khanh P Nguyen

Acknowledgements

Earning a Ph.D. is a long journey and I could not have accomplished this without support from the people around me. I want to take a moment to give special thanks to my supporters.

First, my sincerest thanks go to my advisors, Dr. Krešmir Josić and Dr. Zachary Kilpatrick, who provided me an opportunity to join their lab. Their constant support and faith helped widen my research perspectives, and thus I grew to be a better scientist.

My Ph.D. years would have been painful without my colleagues. I want to thank everyone in Dr. Josić's group, especially Bhargav Karamched, who gave me a necessary push when I was trying to figure out my future career.

I thank Ms. Kathleen Murrell and Ms. Jeanne Parlangeli. They have been my major supporters since high school.

My family plays an important role in my success. I want to thank my parents, Trong & Mai Nguyen, and my only sister, Vi Nguyen, for always supporting me even though they never understood my work. I am grateful to my deceased grandfather, Linh D. Nguyen, for granting me the opportunity to immigrate to the United States and achieve the American dream.

Last but not least, I want to acknowledge the Barry Goldwater Scholarship Foundation and the National Science Foundation for supporting me as a young scientist.

In Loving Memory of my Grandfather: Linh Duc Nguyen

Abstract

To make the best decisions, organisms must flexibly accumulate information, accounting for what is relevant and ignoring what is not. Many decision-making studies focus on sequences of independent trials in which the evidence gathered to make a choice, as well as the resulting actions and feedback, are irrelevant to future decisions. Two-alternative forced choice tasks (2AFC) are often used to characterize strategies subjects use to make decisions. Normative theories, which model ideal observers, have been developed for such tasks when rewards provide the sole evidence (e.g., two-armed bandit tasks). Less is known about how observers should integrate probabilistic rewards interspersed with noisy evidence to inform their decisions in future correlated trials. To understand decision-making under more natural conditions, we extend drift-diffusion models to obtain the normative form of evidence accumulation in a series of 2AFC trials with the correct choice evolving as a two-state Markov process. We analyze three different feedback cases: withholding trial-to-trial feedback, giving probabilistic trial-to-trial signal, and giving probabilistic trial-to-trial reward. Ideal observers integrate noisy evidence within a trial until reaching a decision threshold and bias their initial belief depending on the evidence accumulated and feedback received on previous trials. Optimal observers accumulate more evidence on early trials and make faster decisions on later trials. Gains in performance are primarily due to biases in initial beliefs that lead to faster decisions even when feedback is lacking. Feedback shapes trial-to-trial decision strategies determining whether decisions are immediate, or a result of past and present evidence, depending on whether the

feedback is strong enough to overcome the volatility of changes between trials. Our findings are also consistent with experimentally observed response trends, showing decreased reaction times when correct choices are repeated and in response to prior trial rewards.

Contents

ACKNOWLEDGMENTS	iii
ABSTRACT	v
LIST OF FIGURES	ix
1 Introduction	1
1.1 Evidence of History-Dependent Biases in Decisions	3
1.2 Importance of Feedback in Decision-Making Study	5
1.3 Different Decision-Making Models	8
1.4 Overview	11
2 Optimizing Decisions in the Standard Drift-Diffusion Model	13
2.1 Derivation of the Drift-Diffusion Model for a Single Trial	15
2.2 Mean First Passage Time	17
2.3 Probability of Exit Through a Boundary	19
2.4 The Speed-Accuracy Tradeoff	21
2.4.1 Optimizing Reward Rate	21
2.4.2 Optimizing Expected Net Reward	24
2.5 Summary	24
3 Integrating Information Across Two Correlated Trials	26
3.1 Optimal Observer's Priors	29
3.2 Performance on the First Two Trials	31
3.2.1 Threshold Determines the Probability of Being Correct	32
3.2.2 Mean Response Time on the Second Trial	33
3.3 Performance for Dynamic Decision Thresholds	37
3.4 Optimizing the Expected Net Reward	39
3.5 Summary	42

CONTENTS

4	Integrating Information Across Multiple Correlated Trials	43
4.1	Optimal Observer Model for Multiple Trials	44
4.2	Performance for Constant Decision Thresholds	45
4.3	Performance for Dynamic Decision Thresholds	47
4.4	Performance for Dynamic Decision Thresholds Using Expected Net Reward	49
4.5	Summary	51
5	Effects of Probabilistic Feedback Across Correlated Trials	52
5.1	Feedback Enhances the Observer's Initial Belief	55
5.2	Probabilistic Feedback Reduces Mean Decision Time and Increases Probability of Correct Responses on Trial 2	60
5.2.1	Probability of Being Correct Depends on Reliability of Feed- back	61
5.2.2	Feedback Decreases Mean Decision Time in an Optimal Ob- server	65
5.3	Optimizing Expected Net Reward for Probabilistic Signaling and Probabilistic Reward	68
5.3.1	Probabilistic Signaling	68
5.3.2	Probabilistic Reward	72
5.4	Summary	75
6	Discussion	77
6.1	The Optimal Observer Assumes Incorrect Change Probability . . .	77
6.2	Repetition-Dependence of Correct Probabilities and Decision Times	80
6.3	Our Model Is Consistent With Experimental Trend	83
7	Conclusion & Future Work	88
A	Numerical Methods	95
	Bibliography	96

List of Figures

1.0.1	Random dot motion discrimination task is commonly used to study decision-making processes. Some dots move to one direction, either left or right, while other dots move randomly. Humans or animals need to determine the “correct” direction of the dots. In this experiment, the dot direction on trial 1 is opposite to the ones on trial N.	2
1.1.1	A. There are two possible stimuli that could be presented each trial which the subject must identify. Between trials, the identity of the stimulus is repeated (alternated) with probability $1 - P_A$ (P_A). B. The mean reaction time and the error rate of 16 participants were measured in Goldfarb’s experiment. <i>R</i> stands for Repetition, and <i>A</i> stands for Alternation. Thus, <i>RR</i> , <i>AR</i> , <i>RA</i> , and <i>AA</i> are second-order sequences for repeating and alternating stimuli. [40].	5
1.2.1	Subjects from age groups between 8-9, 11-13, and 18-25 participated in a 2AFC task, where they were instructed to make decisions based on feedback. Positive feedback and negative feedback indicate correct and incorrect choices, respectively. A. Percentage correct. B. Mean reaction time. Overall, subjects perform better when given positive feedback. Ages 8-9 perform better with positive feedback while ages 18-25 can adjust their performance when given negative feedback. The trends are reproduced from Van Duijvenvoorde et al. data [92].	7

2.0.1	A. In a standard two-alternative forced choice task, observers make a sequence of observations, $\xi_{1:n}$, to determine which of the two choices is correct. B. Decision time measures how long it takes $y(t)$ to reach a threshold. Positive/ Negative threshold indicates decisions H_+ / H_- . The correct probability is the probability that $y(t)$ reaches the correct threshold. For example, when $g > 0$, the correct probability is the probability of exiting through the positive threshold when starting at an initial position $y(0)$	14
2.4.1	The variance is $D = 1$. A. Correct probability and B. decision time as functions of decision thresholds θ . As θ increases, it takes more time for the observer to make decisions and the correct probability is improved. C. Reward rate (RR) and D. expected net reward (ENR) as functions of thresholds θ , where the “cost” = 1. The delay time T_D has an impact on the optimum of RR. In contrast, because of the linearity, T_D only causes a vertical shift in ENR.	23
3.0.1	Sequence of 2AFC 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, $H^{1:4} = (H_+, H_+, H_-, H_+)$ stands for the true states of the environment. When the environment is H_+ (H_-), the drift direction of the drift-diffusion process is positive (negative). The decision threshold (θ_{n-1}) crossed on trial $n - 1$ determines the sign of the initial condition, $y^n(0)$, on trial n	27
3.2.1	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 (dots) are compared to theoretical predictions (solid lines). A. The probability of a correct response on trials 1 and 2 ($c_{1:2}$) only depends on the threshold θ , and increases with θ . B. Decision time DT_2 on 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. (3.2.1) 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 time on trial 1 because 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 $D = 1$	36

- 3.3.1 **A,B.** Reward rate (RR), given by Eq. (3.3.3), 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 (dot) satisfies $\theta_1^{\max} > \theta_2^{\max}$. Compare to the pair $\theta_{1:2}^{\max}$ (circle) for $\theta = \theta_{1:2}$ as in Fig. 3.2.1D. The boundary $\theta_2 = y_0^2(\theta_1)$, obtained from Eq. (3.1.4), is shown as a dashed line. For $\epsilon = 0.1$ (panel **B**), the optimal choice of thresholds (dot) is on the boundary $\theta_2 = y_0^2(\theta_1)$. **C.** Decreasing ϵ from 0.5, initially results in an increase in $\theta_1^{\max}$ and decrease in $\theta_2^{\max}$. At sufficiently small value of ϵ , the second threshold collides with the initial condition, $\theta_2 = y_0^2(\theta_1)$ (vertical line). After this point, both $\theta_1^{\max}$ and $\theta_2^{\max}$ increase with ϵ and meet at $\epsilon = 0$. The optimal constant threshold, $\theta^{\max}$, is represented by the thin curve. **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 light curve shows $RR_{1:2}^{\max}$ for the constant threshold case, and the horizontal line shows $RR_{1:2}^{\max}$ when assuming independence. 38
- 3.4.1 Diffusion term $D = 1$, the evidence-accumulation cost is “cost” = 0.5, the intertrial delay time is $T_D = 0.5$, and discount rate $\gamma = 0.7$. **A,B.** Expected Net Reward (ENR), given by Eq. (3.4.1) as a function of change rate ϵ and thresholds $\theta_{1:2}$. The pair $\theta_{1:2}^{\max}$ that maximizes ENR (dot) satisfies $\theta_1^{\max} > \theta_2^{\max}$. On the other hand, the pair $\theta_{1:2}^{\max}$ (circle) maximizes ENR under the constraint on constant thresholds, $\theta_1 = \theta_2$. The boundary $\theta_2 = y_0^2(\theta_1)$, obtained from Eq. (3.1.4), is shown as a dashed line. **C, D.** Optimal decision thresholds $\theta_{1:2}^{\max}$ and maximum $ENR_{1:2}^{\max}$ as functions of change rate ϵ . The trends are similar to the results in Fig. 3.3.1 **C** and **D** even though the optimal values are different between RR and ENR. The thin light blue curve shows $ENR_{1:2}^{\max}$ for the constant threshold case, and the vertical line in panel **C** shows the boundary of $\theta_2 = y_0^2(\theta_1)$ 41
- 4.2.1 Reward rate ($RR_{1:n}$) defined by Eq. (4.2.2) 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, $\theta^{\max}$, increases with n . The limiting reward rates $RR_{1:n}(\theta; \epsilon) \rightarrow RR_{\infty}(\theta; \epsilon)$ as $n \rightarrow \infty$ defined by Eq. (4.2.3) are given by the topmost light curves. 46

- 4.3.1 Optimal sets of thresholds, $\theta_{1:n}^{\max}$, for n trials with correlated states. Top row: **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**). Thin curves in each plot show the optimal choice of constant threshold θ . Bottom row: Fixing **D**. $\epsilon = 0.01$, **E**. $\epsilon = 0.1$, and **F**. $\epsilon = 0.25$, the optimal threshold vectors $\theta_{1:n}^{\max}$ are displayed as a function of trial number. The stereotyped decrease in $\theta_j^{\max}$ is preserved as a function of j across different total trial numbers n . The decrease is more gradual for small ϵ , whereas for larger ϵ , thresholds are constant until the last trial $\theta_n^{\max}$ 49
- 4.4.1 Optimal sets of thresholds computed from ENR, $\theta_{1:n}^{\max}$, for n trials with correlated states. Here, $D = 1$, $\text{cost} = T_D = 0.5$ and $\gamma = 0.7$. **A**. For $n = 3$, the set of optimal thresholds $\theta_{1:3}^{\max}$ still 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**. For $n = 5$, the ordering of optimal thresholds is consistent, i.e., higher thresholds on earlier trials. Thin curves in each plot show the optimal choice of constant threshold θ 50
- 5.0.1 True environmental state $H^n \in H_{\pm}$ switches stochastically between trials with probability ϵ . An observer sequentially accumulates noisy evidence within a trial and biases their initial belief based on the past choice and feedback. When the green light (or reward) appears at the end of trial 1, the initial belief on trial 2 may end up above the decision boundary due to the additional positive information. Thus, the observer may use feedback at the end of trial 1 to make an immediate decision on trial 2. **A**. Given probabilistic signal as feedback, despite the erroneous red (negative) signal at the end of trial 2, the observer will still receive a reward for trial at the end of the trial sequence, since they were correct. **B**. In the case of probabilistic reward, the observer is sometimes not rewarded/rewarded even though they are correct/incorrect. In this example, there is no reward on trial 2 even though the observer is correct. 53

- 5.1.1 Color plots of initial belief on trial 2, $y_{0\pm}^2$, given by light signals: **A.** green light, **B.** red light; and by reward feedback: **C.** reward, **D.** no reward. The optimal policy is for constant thresholds across trials. The curves indicate the parameter values (ϵ, p) at which $\theta_s^{\max} = |y_{0\pm}^2|$, or $\theta_r^{\max} = |y_{0\pm}^2|$, i.e., decisions occur immediately. The solid (dashed) curve corresponds to the case of probabilistic signaling (probabilistic reward) while the green (blue) color indicates positive (negative) feedback. 59
- 5.2.1 Here, the probability of environmental change is $\epsilon = 0.25$ and probability of correct feedback is $p = 0.7$. The stochastic simulations (dots) are compared to theoretical predictions (solid lines). Feedback improves the probability of being correct in the case of quicker decisions. Positive feedback (green curves) leads to both more accurate and speedier decisions as compared to decisions following negative feedback (red curves). This is because most positive feedback follows a correct decision. **A.** Probability of correct responses. **B.** Mean decision time on trial 2. 64
- 5.3.1 In the case of probabilistic signaling, color plots of **A.** Optimal threshold θ and **B.** the maximum ENR. The green (blue) curve denotes where $|y_{0+}^2| = \theta_s^{\max}$ ($|y_{0-}^2| = \theta_s^{\max}$), so in the region (ϵ, p) to the left of these curves, responses are immediate on trial 2. 69
- 5.3.2 Color plots of probability of correct response and mean decision time on trial 2 in the context of probabilistic signaling with parameter p . The green (blue) solid curve indicates where $|y_{0+}^2| = \theta_s^{\max}$ ($|y_{0-}^2| = \theta_s^{\max}$). Optimal decision thresholds $\theta_s^{\max}$ are found by optimizing Eq. (5.3.1). The regions above the curves are “immediate-decision” regions. Correct probability on trial 2 conditioned on **A.** green light, $c_2|(s_1 = +1)$, **B.** red light, $c_2|(s_1 = -1)$, and **C.** probability being correct c_2 . **D.** Decision time on trial 2 conditioned on **D.** green light, $DT_2|(s_1 = +1)$; **E.** conditioned on red light, $DT_2|(s_1 = -1)$, and **F.** mean decision time on trial 2, DT_2 71
- 5.3.3 In the case of probabilistic reward, color plots of **A.** optimal threshold $\theta_r^{\max}$ and **B.** maximum $ENR_{1:2}^{\max}$. The green (blue) dashed curve is $|y_{0+}^2| = \theta_r^{\max}$ ($|y_{0-}^2| = \theta_r^{\max}$). Note that the regions in which an instantaneous decision is triggered on trial 2 (left of curves) is larger than in the case of probabilistic signaling. 73

- 5.3.4 **A, B, C.** Color plots of probability of correct response and **D, E, F.** mean decision time on trial 2 following a probabilistic reward on trial 1. The green (blue) curve indicates the collision between initial belief for positive (negative) feedback and threshold $y_{0+}^2 = \theta_r^{\max}$ ($y_{0-}^2 = \theta_r^{\max}$). The positive (negative) feedback is defined as the observer receive reward (no reward). Optimal thresholds $\theta_r^{\max}$ are found from optimizing Eq. (5.3.2). **A, D.** The conditional correct probability and decision time given by positive feedback. **B, E.** The conditional correct probability and decision time given by negative feedback. **C, F.** Total correct response and mean decision time. . . . 74
- 6.1.1 The optimal RR for a two trial task as a function of the true ϵ_{true} and the assumed $\epsilon_{\text{assumed}}$ switching rates. Eq. (3.2.1) is maximized by tuning (θ_1, θ_2) while fixing $(\epsilon_{\text{true}}, \epsilon_{\text{assumed}})$. **A.** For slow environments (small ϵ_{true}), the RR obtains the same maximum regardless of the choice of $\epsilon_{\text{assumed}} < 0.5$, while for larger ϵ_{true} the optimal choice $\epsilon_{\text{assumed}}^{\max} = \epsilon_{\text{true}}$. **B.** The RR varies weakly with $\epsilon_{\text{assumed}}$ about the maximum RR (dots) for each ϵ_{true} 78
- 6.3.1 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 (top curve). When $d_{j-1} = -1$, the fraction of $d_j = +1$ responses decreases (bottom curve). The weighted average of these two curves gives the unconditioned psychometric function (middle curve). Coherence g/D is defined for drift $g \in \pm 1$, noise diffusion coefficient D , and assumed change rate $\epsilon_{\text{assumed}} = 0.25$. 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 $\epsilon_{\text{assumed}}$, 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 $\epsilon_{\text{assumed}}$ is increased, since the effects of the state two trials back are weakened. Compare dots at $\epsilon_{\text{assumed}} = 0.05$ and $\epsilon_{\text{assumed}} = 0.25$. **C.** Decision time DT_{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 $\epsilon_{\text{assumed}} = 0.1, 0.25, 0.4$. In all cases, we set $T_D = 2$, $\theta = \theta_{\infty}^{\max}$ for all trials, and $D = 1$ if not mentioned. 85

Introduction

Organismal behavior is often driven by decisions that are the result of evidence accumulated to determine the best choice among available options [12, 39]. Often such evidence comes from conspecifics: 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 [82]. Competitive animals evaluate their opponents’ attributes to decide whether to fight or flee [85], and humans decide which stocks to buy or sell, based on individual research and social information [57]. Importantly, the observations of these agents are frequently uncertain [15, 42] so accurate decision-making requires robust evidence integration that accounts for the reliability and variety of evidence sources [70].

The two-alternative forced choice (2AFC) task paradigm has been successful in probing the behavioral trends and neural and cognitive mechanisms underlying decision-making [71]. In a 2AFC task, subjects decide which one of two hypotheses is more likely based on noisy evidence [8, 38]. For instance, in a random dot

motion discrimination task (see Fig. 1.0.1), 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, showing that the oculomotor system either takes part or is driven by the evolving decision-making process. This observation also motivates the use of such stimuli in the experimental study of neural mechanisms underlying decision-making [39, 83].

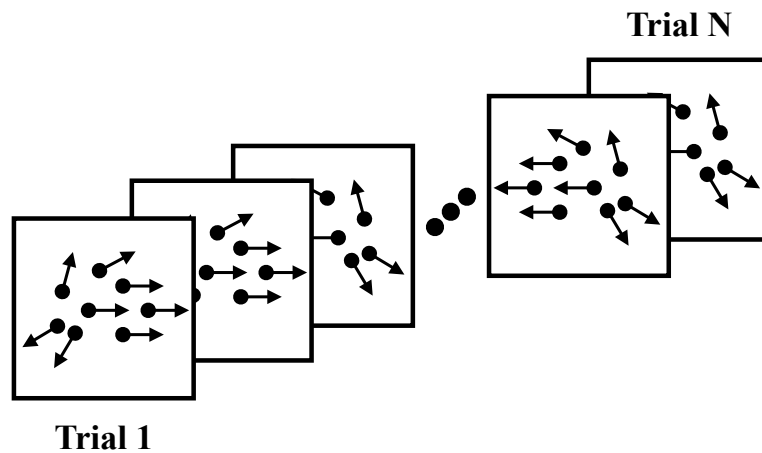


Fig. 1.0.1: Random dot motion discrimination task is commonly used to study decision-making processes. Some dots move to one direction, either left or right, while other dots move randomly. Humans or animals need to determine the “correct” direction of the dots. In this experiment, the dot direction on trial 1 is opposite to the ones on trial N.

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 crosses [8, 73, 86].

1.1 Evidence of History-Dependent Biases in Decisions

To survive, animals must constantly learn about and adapt to the changing environment. Some changes are a result of autonomous cycles in the environment (e.g., temperature, seasons), while others are due to the actions of living beings. For example, a group of animals migrating can cause food shortages for other animals, or a new predator in the area can influence prey's foraging decisions [97]. On the other hand, humans often make decisions to suit their needs. For example, if you enjoyed a dish you ordered on your last visit to a restaurant, you are likely to order the same dish at your next visit [43]. Experiments have shown that both humans and animals can exhibit several irrational trends in their choices based on previous ones. Such trends are often referred to as sequential effects or biases [17, 84].

Soetens et al. conducted experiments demonstrating sequential effects where human subjects decided if a LED stimulus appeared on the left or right [84]. Subjects' mean reaction time and error rate were recorded according to XXXX stimulus sequences (the previous four stimulus sides relative to the current one), where X was R (A), which stood for a stimulus repetition (alternation), given 16 total possible sequences. The data showed that subjects react faster in sequences of AAAA type compared to sequences of AAAR type. Intuitively, subjects are likely to expect the stimulus to alternate after three consecutive alternations. They thus

react faster when this expectation is met, and slower when surprised by a repeating stimulus. Similar trends were also observed for other types of stimulus patterns [84]. Humans use accumulated information across trials to bias their belief in current trials, which can improve their performance if the strategy is well-matched to environmental trends but will decrease performance if the strategy is mismatched.

History-dependent biases are also found in animals. Such biases can also be reflected in neural activity: Hwang et al. examined the pre-trial activity of the posterior parietal cortex (PPC) in mice. In this task, mice decided if a visual stimulus was forward or downward. A significant choice selectivity was found in a large fraction of choice-selective neurons in PPC. Since PPC neurons encode the history of outcomes and transform visual information to motor outputs, the result implies that neurons had predicted the choice before the trial even started [43]. In a recent study of primates' behaviors in an auditory task, in which monkeys decided if the frequency of a sequence of tone bursts was predominantly low or high, it was found that choice selectivity increases in ventrolateral prefrontal cortex (vIPFC) neurons post-decision and pre-reward. The vIPFC neurons linked auditory perception and decision-making, so activation in this area implies that biases on subsequent trials are established as soon as the current decision is committed [89]. Both studies show that animals form biases at the beginning of each subsequent trial.

In these experiments, stimuli appeared randomly on each trial and there were no intertrial correlations. The observed biases imply that humans and animals

1.2. IMPORTANCE OF FEEDBACK IN DECISION-MAKING STUDY

could be assuming that trials are correlated, even when no correlations exist. Goldfarb et al. performed related experiments examining responses in a series of dependent trials with the correct choice across trials evolving according to a two-state Markov process [40]. The transition probabilities affected the response time and accuracy of subjects in ways well described by a DDM with biased initial conditions and thresholds (see Fig. 1.1.1 A). For instance, when repeated states were more likely, response times decreased in the second of two repeated trials (see Fig. 1.1.1 B).

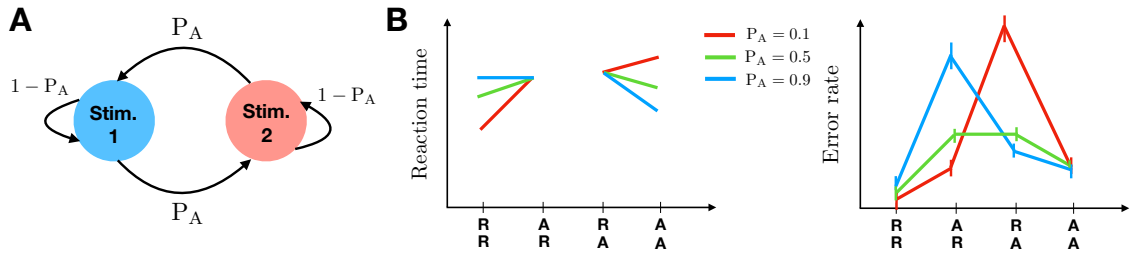


Fig. 1.1.1: **A.** There are two possible stimuli that could be presented each trial which the subject must identify. Between trials, the identity of the stimulus is repeated (alternated) with probability $1 - P_A$ (P_A). **B.** The mean reaction time and the error rate of 16 participants were measured in Goldfarb's experiment. *R* stands for Repetition, and *A* stands for Alternation. Thus, *RR*, *AR*, *RA*, and *AA* are second-order sequences for repeating and alternating stimuli. [40].

1.2 Importance of Feedback in Decision-Making Study

In nature, humans and animals require feedback to improve their future performance. For example, both songbirds and humans rely on auditory feedback to learn a dialect [10]. Toddlers receive feedback to learn simple tasks while adults

1.2. IMPORTANCE OF FEEDBACK IN DECISION-MAKING STUDY

need feedback to learn more complicated tasks [66, 92]. A student needs feedback to learn how to play a violin [53]. Because our world is uncertain, sometimes feedback is not perfect. We are interested in how an optimal observer combines probabilistic feedback and their internal belief to make decisions in a changing environment.

In experiments, one way to give subjects feedback is to give a reward at the end of each trial. This type of feedback is more common in animal studies. The effect of feedback has been an interesting topic in the decision-making field. The neural activity of dopamine neurons is typically examined to study this effect because dopamine plays an important role in motivating action [13, 14, 96]. Moreover, the mismatch between expected and actual reward is coded by dopamine neurons. A surprising reward (or omission of a reward) activates (or depresses) dopamine neurons [80, 94]. In a high-variance environment, subjects who receive more frequent feedback, and in this case, we can interpret that as receiving more frequent rewards, perform worse [55]. Therefore, it is important for us to explore and quantify the optimal strategy when the feedback is probabilistic.

On the other hand, human subjects are able to adjust their strategies based on true/false feedback. Duijvenvoorde et al. examined the performance and neural activity of subjects across a range of different ages while they perform feedback-based sequential tasks [92]. Positive feedback encourages subjects to make the same decision on repeated trials, while negative feedback motivates subjects to switch their choice. Experimental data showed that accuracy increased and reaction time decreased across age groups for positive feedback (see Fig. 1.2.1) while

1.2. IMPORTANCE OF FEEDBACK IN DECISION-MAKING STUDY

only adults successfully adjusted their behaviors after negative feedback. This result was consistent with previous studies [19, 79]. Moreover, more activation was found to the left of the dorsolateral prefrontal cortex (DLPC) in adults post-negative feedback. This contrasted with what was found in children, as there was more activation post-positive feedback. In summary, both positive and negative feedback play important roles in sequential tasks.

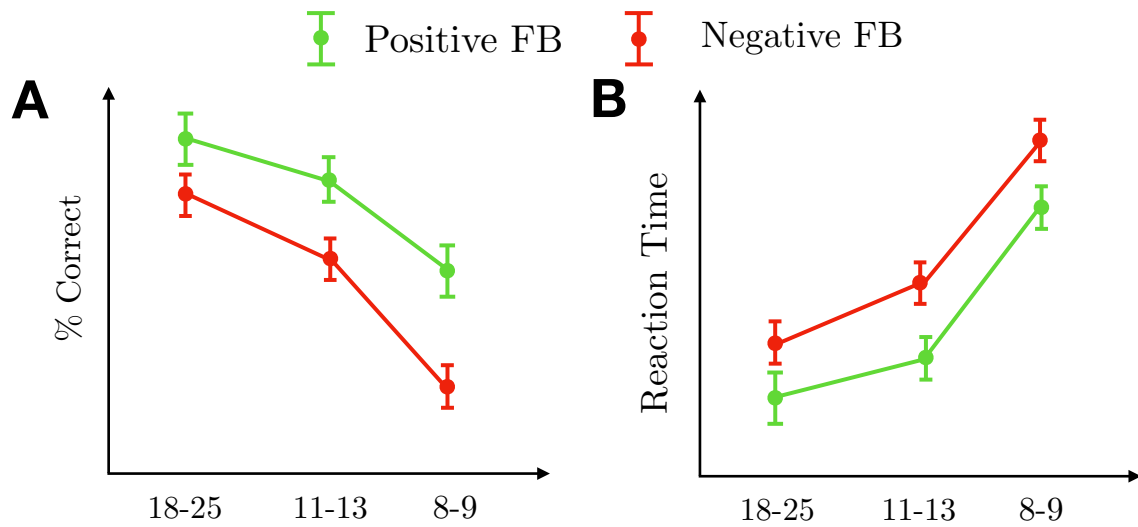


Fig. 1.2.1: Subjects from age groups between 8-9, 11-13, and 18-25 participated in a 2AFC task, where they were instructed to make decisions based on feedback. Positive feedback and negative feedback indicate correct and incorrect choices, respectively. **A.** Percentage correct. **B.** Mean reaction time. Overall, subjects perform better when given positive feedback. Ages 8-9 perform better with positive feedback while ages 18-25 can adjust their performance when given negative feedback. The trends are reproduced from Van Duijvenvoorde et al. data [92].

1.3 Different Decision-Making Models

Even though there are many mathematical decision-making models (such as leaky competing accumulator, integration to bound, urgency model, etc.), the drift-diffusion model (DDM) is typically used to study this decision-making process. The DDM is popular because (1) it can be derived as the continuum limit of the statistically-optimal sequential probability ratio test [8, 95]; (2) it is an analytically tractable Wiener diffusion process whose summary statistics can be computed explicitly [8, 74]; and (3) it can be fit remarkably well to behavioral responses and neural activity in 2AFC tasks with independent trials [38, 39] (although see Latimer et al. [52]).

However, 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 [73]. Organisms in nature often make a sequence of related decisions based on overlapping evidence [16]. 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 [18]. Animals foraging over multiple days are biased towards food sites with consistently high yields [41]. Thus even in a variable environment, organisms use previously gathered information to make future decisions [20].

Even in a sequence of independent trials, previous choices influence subsequent decisions [32]. Such serial response dependencies or *recency effects* have

been observed in 2AFC tasks which do [4, 17, 34] and do not [6, 50] 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 [65]. Thus, trends in response time and accuracy suggest subjects use trial history to predict the current state of the environment, albeit suboptimally [50].

History-dependent biases also increase the probability of repeat responses when subjects view sequences with repetition probabilities above chance [1, 11]. 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 [40]. There is an interesting proposal that bias can be due to a drift shift. The choice pattern of initial condition bias and drift bias are identical with different reaction time distributions [90].

Recency effects can appear if an observer is learning the statistical structure of the environment. Much of the early literature on this subject treated recency biases as a consequence of an observer trying to learn the frequency of events in a sequence (reviewed in [31]). For instance, classical incremental learning [27] has been used to explain how learning the probability of each of two choices across trials can lead to responses biased by recent feedback [26].

Learning trial-to-trial dependencies can also lead to recency effects [17, 30, 97]. In combination with learning the marginal frequency of each choice, this can explain a range of experimental observations [45, 49]. However, there is substantial evidence that human observers can learn dependencies between different trials

[4], and that recency effects are the result of observers stably learning such structure [47, 58]. Moreover, to address the reward prediction for dopamine neuron output in primates, temporal difference (TD) method is applied because (1) it predicts a discounted sum of future rewards using sensory cues; (2) it assumes Markovian property [81]. In particular, as we will discuss further in this work, evidence accumulated on a previous trial and feedback can bias an observer’s initial belief and their response on the next trial.

There are many non-DDM models that describe sequential effects [4, 17, 27, 51]. For instance, Cho et al. used Usher and McClelland’s leaky competing accumulator model to study this effect in 2AFC tasks by constraining possibilities of bias updates [17]. Even though this model reproduces behavioral trends better than previous models, it only induced a sub-optimal strategy. Yu et al. proposed the *Fixed Belief Model* using the Bayesian framework to study sequential decision-making. The model shows that in a changing environment, humans weigh past observations linearly with an exponential decaying coefficient. The optimal Bayesian inference is approximated by this discounting, but error patterns were not accounted for [97]. However, these models are not normative and we are not aware of previous normative models that describe both recency and feedback effects in evidence accumulation tasks. Normative models are useful because they describe the behaviors of ideal observers.

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 and feedback reliability are known to the observer. We use sequential analysis to derive the posterior for the environmental states $H_{\pm}$ given a stream of noisy observations [8, 93]. 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 a 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 show that even in the absence of fractional learning, i.e., when the observer is fully aware of the correlations between trials, a temporal structure of responses similar to that observed experimentally can emerge due to the true or assumed dependencies between trials.

1.4 Overview

Many 2AFC tasks have been designed so that the stimulus itself has no correlations from one trial to the next, and often modelers do not consider the effects of reward or feedback on decision strategy [39]. In this dissertation, we focus on two aspects of a more realistic decision-making model: correlations between trials and probabilistic feedback. In chapter 1, we introduce experiments that motivate our study. In chapter 2, we re-derive the standard drift-diffusion model (DDM) when sequences of trials are uncorrelated, its summary statistics, as well as expressions for first passage times and exit probabilities. We will use these computations in later chapters to compute *reward rate* and *expected net reward* functions, which are measurements of the optimal observer's performance. In chapter 3, we consider

the first two trials of the sequence. We assume that the correlation is known to an optimal observer and feedback is omitted. When the correlation is strong, the observer carries more information on subsequent trials to improve the reaction time. We derive initial bias and compute the observer's performance in cases of constant and dynamic decision thresholds. We extend the findings to multiple trials in chapter 4. In chapter 5, we consider optimal performance in correlated trials with the observer receiving noisy feedback. Such feedback can take two forms: (1) a probabilistic light signal with a deterministic reward, or (2) a probabilistic reward. We recap our findings in chapter 6 by comparing our model to experimental data response trends.

Optimizing Decisions in the Standard Drift-Diffusion Model

The standard DDM can be derived as the continuum limit of a recursive equation for the log-likelihood ratio (LLR) [8, 38, 95]. When the environmental states are uncorrelated between trials and equally likely, an ideal observer has no bias at the start of each trial. We assume that on one trial the observer integrates a stream of noisy measurements of the true state, H (see Fig. 2.0.1 A). If these measurements are conditionally independent, the functional central limit theorem yields the DDM for the scaled LLR, $y(t) = D \ln \frac{P(H=H_+|\text{observations})}{P(H=H_-|\text{observations})}$, after observation time t ,

$$dy = gdt + \sqrt{2D}dW. \tag{2.0.1}$$

We have written the above as a Langevin equation in which W is a Wiener

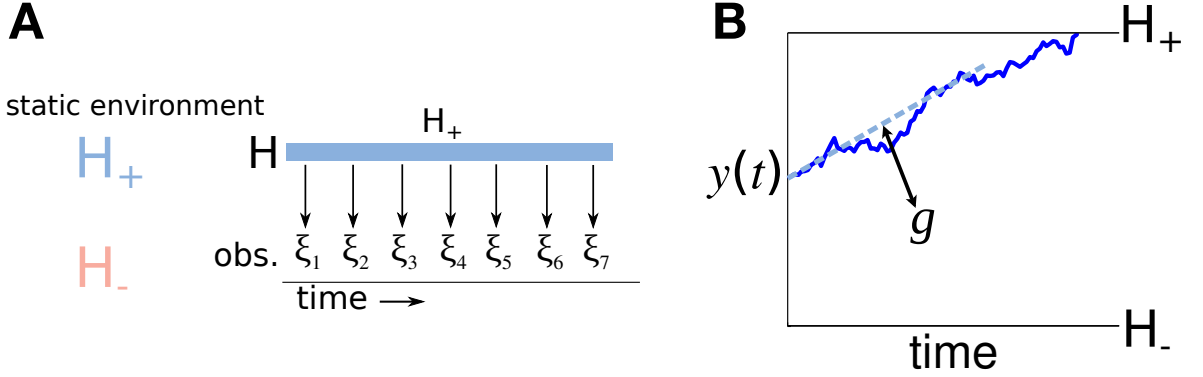


Fig. 2.0.1: **A.** In a standard two-alternative forced choice task, observers make a sequence of observations, $\xi_{1:n}$, to determine which of the two choices is correct. **B.** Decision time measures how long it takes $y(t)$ to reach a threshold. Positive/Negative threshold indicates decisions H_+/H_- . The correct probability is the probability that $y(t)$ reaches the correct threshold. For example, when $g > 0$, the correct probability is the probability of exiting through the positive threshold when starting at an initial position $y(0)$.

process, the drift $g \in g_{\pm}$ depends on the environmental state, and D is the variance which depends on the quality of each observation. Often, the prefactor of the Weiner process is taken to be σ , rather than $\sqrt{2D}$, but the form we have chosen simplifies several expressions we derive subsequently. Eq. (2.0.1) may be written as a stochastic differential equation $\dot{y} = g + \sqrt{2D}\zeta$, where ζ is Gaussian white noise, and some use the form $dy = gdt + \zeta \cdot \sqrt{2D \cdot dt}$ [91].

For simplicity and consistency with typical random dot kinetogram tasks [39, 78], we assume each drift direction is of equal unit strength: $g_+ = -g_- = 1$, and task difficulty is controlled by scaling the variance through D . An 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 initial condition is

2.1. DERIVATION OF THE DRIFT-DIFFUSION MODEL FOR A SINGLE TRIAL

determined by the observer's prior bias, $y(0) = D \ln[P(H = H_+)/P(H = H_-)]$. Thus for an unbiased observer, $y(0) = 0$.

There are two primary ways of obtaining a response from the DDM given by Eq. (2.0.1) that mirror common experimental protocols [38, 39, 83]. The first way is that the ideal observer interrogates at a set time, $t = T$, responds with $\text{sign}(y(T)) = \pm 1$ indicating the more likely of the two states, $H = H_{\pm}$, 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 , and assuming that at the first time, T , at which $|y(T)| \geq \theta$, the evidence accumulation process terminates, and the observer chooses $H_{\pm}$ if $\text{sign}(y(T)) = \pm 1$ (see Fig. 2.0.1 B).

2.1 Derivation of the Drift-Diffusion Model for a Single Trial

We assume that the optimal observer integrates a stream of noisy measurements $\xi_{1:m} = (\xi_1, \xi_2, \dots, \xi_m)$ at equally spaced times $t_{1:m} = (t_1, t_2, \dots, t_m)$ [8, 95]. 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 on 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

$$R_m = \frac{P(H = H_+|\xi_{1:m})}{P(H = H_-|\xi_{1:m})} = \frac{f_+(\xi_1)f_+(\xi_2) \cdots f_+(\xi_m)P(H = H_+)}{f_-(\xi_1)f_-(\xi_2) \cdots f_-(\xi_m)P(H = H_-)}, \quad (2.1.1)$$

2.1. DERIVATION OF THE DRIFT-DIFFUSION MODEL FOR A SINGLE TRIAL

due to the independence of each measurement ξ_m . Thus, if $R_m > 1$ ($R_m < 1$) then $H = H_+$ ($H = H_-$) is the more likely state. Eq. (2.1.1) can be written recursively

$$R_m = \left(\frac{f_+(\xi_m)}{f_-(\xi_m)} \right) R_{m-1}, \quad (2.1.2)$$

where $R_0 = P(H_+)/P(H_-) = 1$ due to the equal probability of both states at the beginning of the first trial.

Taking the logarithm of Eq. (2.1.2) allows us to express the recursive relation for the log-likelihood ratio (LLR) $L_m^1 = \ln R_m$ as an iterative sum

$$L_m = L_{m-1} + \ln \frac{f_+(\xi_m)}{f_-(\xi_m)},$$

where $L_0 = \ln [P(H_+)/P(H_-)]$, so if $L_m \gtrless 0$ then $H = H_{\pm}$ is the more likely state. Taking the continuum limit $\Delta t \rightarrow 0$ of the timestep $\Delta t := t_m - t_{m-1}$, we can use the functional central limit theorem to yield the DDM (see [8, 93] for details)

$$dy = gdt + \sqrt{2D}dW, \quad (2.1.3)$$

where W is a Wiener process, the drift $g \in g_{\pm} = \frac{1}{\Delta t} E_{\xi} \left[\ln \frac{f_+(\xi)}{f_-(\xi)} | H_{\pm} \right]$ 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. (2.0.1) in hand, we can relate $y(t)$ to the probability of either state $H_{\pm}$

by noting its Gaussian statistics in the case of free boundaries

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

so that

$$\frac{P(H = H_+|y(t))}{P(H = H_-|y(t))} = \frac{P(y(t)|H = H_+)}{P(y(t)|H = H_-)} = e^{y(t)/D}. \quad (2.1.4)$$

Note that an identical relation was obtained in [8] in the case of absorbing boundaries. Either way, it is clear that $y = D \cdot \text{LLR}$. This means that before any observations have been made $e^{y(0)/D} = \frac{P(H=H_+)}{P(H=H_-)}$, so we scale the log ratio of the prior by D to yield the initial condition y_0 in Eq. (2.0.1).

2.2 Mean First Passage Time

Here we provide a brief overview of how to compute the mean first passage time in a DDM by analyzing the corresponding Kolmogorov backward equation [36]. Assume the decision variable, y , evolves according to Eq. (2.0.1) with positive drift $g^1 = 1$ (a similar derivation can be carried out for $g^1 = -1$), and let $p(y, t|y_0, 0)$ denote the conditional density of state y at time t given initial condition y_0 at time $t = 0$ whose corresponding Kolmogorov equation is

$$\partial_t p(y, t|y_0, 0) = \partial_{y_0} p(y, t|y_0, 0) + D \cdot \partial_{y_0}^2 p(y, t|y_0, 0), \quad (2.2.1)$$

2.2. MEAN FIRST PASSAGE TIME

given initial condition $p(y, 0|y_0, 0) = \delta(y - y_0)$ and boundary conditions $p(y, t|\pm\theta, 0) = 0$ for $y \in (-\theta, \theta)$, implying absorbing boundaries at $y_0 = \pm\theta$. The survival probability, that $y \in (-\theta, \theta)$ at time t , is defined:

$$G(y_0, t) := \int_{-\theta}^{\theta} p(y, t|y_0, 0) dy = P(T \geq t),$$

where T is a random variable, corresponding to the time the state y leaves the domain $y \in (-\theta, \theta)$. Integrating Eq. (2.2.1), we obtain

$$\partial_t G(y_0, t) = \partial_{y_0} G(y_0, t) + D \cdot \partial_{y_0}^2 G(y_0, t). \quad (2.2.2)$$

Boundary conditions for Eq. (2.2.1) imply $G(\pm\theta, t) = 0$.

The mean first passage time, $T(y_0; \theta)$, for the state to depart the domain $(-\theta, \theta)$ given initial condition y_0 is then given by marginalizing t against the flux $\partial_t G(y_0, t)$ of survival probability out of $(-\theta, \theta)$:

$$T(y_0; \theta) = - \int_0^\infty t \partial_t G(y_0, t) dt = \int_0^\infty G(y_0, t) dt.$$

Noting $\int_0^\infty G(y_0, t) dt = \lim_{t \rightarrow \infty} G(y_0, t) - G(y_0, 0) = -1$ and integrating Eq. (2.2.2), we find

$$\partial_{y_0} T(y_0; \theta) + D \cdot \partial_{y_0}^2 T(y_0; \theta) = -1,$$

with boundary condition $T(\pm\theta; \theta) = 0$, which can be solved explicitly

$$T(y_0; \theta) = \theta \left[\frac{e^{\theta/D} + e^{-\theta/D} - 2e^{-y_0/D}}{e^{\theta/D} - e^{-\theta/D}} \right] - y_0, \quad (2.2.3)$$

which for $y_0 = 0$ simplifies to

$$T(0; \theta) = \theta \left[\frac{1 - e^{-\theta/D}}{1 + e^{-\theta/D}} \right] = DT. \quad (2.2.4)$$

2.3 Probability of Exit Through a Boundary

To derive the probability of exit from either boundary $y = \pm\theta$, when the decision variable begins at y_0 , we integrate the probability current through the boundary of interest, $\pm J(\pm\theta, t|y_0, 0)$. For instance, the probability the decision variable exits $y = +\theta$ after time t is

$$g_\theta(y_0, t) = \int_t^\infty J(\theta, t'|y_0, 0) dt' = \int_t^\infty [p(\theta, t'|y_0, 0) - D\partial_{y_0}p(\theta, t'|y_0, 0)] dt'. \quad (2.3.1)$$

Using the fact that $p(\theta, t|y_0, 0)$ satisfies Eq. (2.2.1), we find that $g_\theta(y_0, t)$ satisfies

$$\partial_{y_0}g_\theta(y_0, t) + D\partial_{y_0}^2g_\theta(y_0, t) = \int_t^\infty \partial_{t'}J(\theta, t'|y_0, 0) dt' = J(\theta, t|y_0, 0) = \partial_tg_\theta(y_0, t). \quad (2.3.2)$$

2.3. PROBABILITY OF EXIT THROUGH A BOUNDARY

Taking $t \rightarrow 0$ and defining $\pi_\theta(y) := g_\theta(y_0, 0)$, we see that $J(\theta, 0|y_0, 0)$ vanishes if $\theta \neq y_0$, since $p(\theta, 0|y_0, 0) = \delta(y_0 - \theta)$, so the right hand side goes to zero and

$$\partial_{y_0} \pi_\theta(y_0) + D \partial_{y_0}^2 \pi_\theta(y_0) = 0, \quad (2.3.3)$$

where $\pi_\theta(\theta) = 1$, $\pi_\theta(-\theta) = 0$, and $\pi_\theta(y_0) + \pi_{-\theta}(y_0) = 1$. Solving the above equations explicitly, we find

$$\pi_\theta(y_0) = \frac{1 - e^{-(y_0+\theta)/D}}{1 - e^{-2\theta/D}}, \quad \pi_{-\theta}(y_0) = \frac{e^{-(y_0+\theta)/D} - e^{-2\theta/D}}{1 - e^{-2\theta/D}}. \quad (2.3.4)$$

An exit through the threshold θ results in a correct choice of $H = H_+$, so $c = \pi_\theta(0)$. The correct probability c increases with θ , since more evidence is required to reach a larger threshold. Defining the decision as $d = \pm 1$ if $y(T) = \pm\theta$, Bayes' rule implies

$$\frac{1}{1 + e^{-\theta/D}} = c = P(d = \pm 1 | H = H_\pm) = P(H = H_\pm | d = \pm 1), \quad (2.3.5a)$$

$$\frac{e^{-\theta/D}}{1 + e^{-\theta/D}} = 1 - c = P(d = \pm 1 | H = H_\mp) = P(H = H_\mp | d = \pm 1), \quad (2.3.5b)$$

since $P(H = H_\pm) = P(d = \pm 1) = 1/2$. Rearranging the expressions in Eq. (2.3.5) and isolating θ relates the threshold θ to the LLR given a decision $d = \pm 1$

$$\pm\theta = D \ln \frac{P(H = H_+ | d = \pm 1)}{P(H = H_- | d = \pm 1)}.$$

2.4 The Speed-Accuracy Tradeoff

Humans and other animals balance speed and accuracy to maximize the rate of correct decisions [9, 16, 39]. If an observer spends more time to accumulate evidence, they improve their accuracy. However, spending too much time on one trial will decrease the performance overall. The value of decision threshold θ reflects this speed-accuracy tradeoff [8]. Setting a high threshold will increase both probability of being correct (see Fig. 2.4.1 A) and decision time (see Fig. (2.4.1 B)). We can measure the observer's performance by computing reward rate (RR) or expected net reward (ENR).

The RR provides a measure of how much reward is gained on average in a given unit of time and is more relevant when considering long trial sequences in which such averages will be realized. Since it is a rate, the time per trial enters as a divisor of the RR. Alternatively, one may consider the ENR in which the observer's time spent deciding penalizes their reward as a linear subtractive term. Due to the linearity of the ENR function, methods from dynamic programming can be applied more readily [23]. However, in some instances, it is possible to rescale the RR function so that a linear treatment in dynamic programming is also possible.

2.4.1 Optimizing Reward Rate

Performance can be quantified as reward rate (RR) function, or total reward per unit time. Earlier decisions are preferred since decreasing reaction time leads to higher RR [23, 38]. The most basic functional expression of the RR is typically

given as

$$\text{RR}(\theta) = \frac{c}{\text{DT} + T_D}, \quad (2.4.1)$$

where c is the probability of a correct decision, DT is the mean time required for $y(t)$ to reach either threshold $\pm\theta$, and T_D is the prescribed time delay to the start of the next trial [8, 38]. Without the delay time, the optimal strategy is to react instantaneously without accumulating evidence because the probability being correct will be 0.5 while the decision time is 0. Adding the term T_D prevents the observer from making decisions randomly, and such delay times are typically implemented in experiments. Eq. (2.4.1) increases with c (accuracy) and with inverse time $1/(\text{DT} + T_D)$ (speed). This usually leads to a nonmonotonic dependence of RR on the threshold, θ , since increasing θ increases accuracy and decreases speed [8, 38] (see Fig. 2.4.1 C).

Substituting this expression into the RR as defined in Eq. (2.4.1), assuming $y_0 = 0$, we have

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

We can identify the maximum of $\text{RR}(\theta) > 0$ by finding the minimum of its reciprocal, $1/\text{RR}(\theta)$ [8],

$$\theta^{\text{opt}} = T_D + D - D\mathcal{W}(\exp[(T_D + D)/D]). \quad (2.4.2)$$

Here $\mathcal{W}(f)$ is the Lambert W function (the inverse of $f(\mathcal{W}) = \mathcal{W}e^{\mathcal{W}}$). In the limit

2.4. THE SPEED-ACCURACY TRADEOFF

$T_D \rightarrow 0$, we have $\theta^{\text{opt}} \rightarrow 0$, and Eq. (2.4.2) defines nontrivial optimal thresholds at $T_D > 0$.

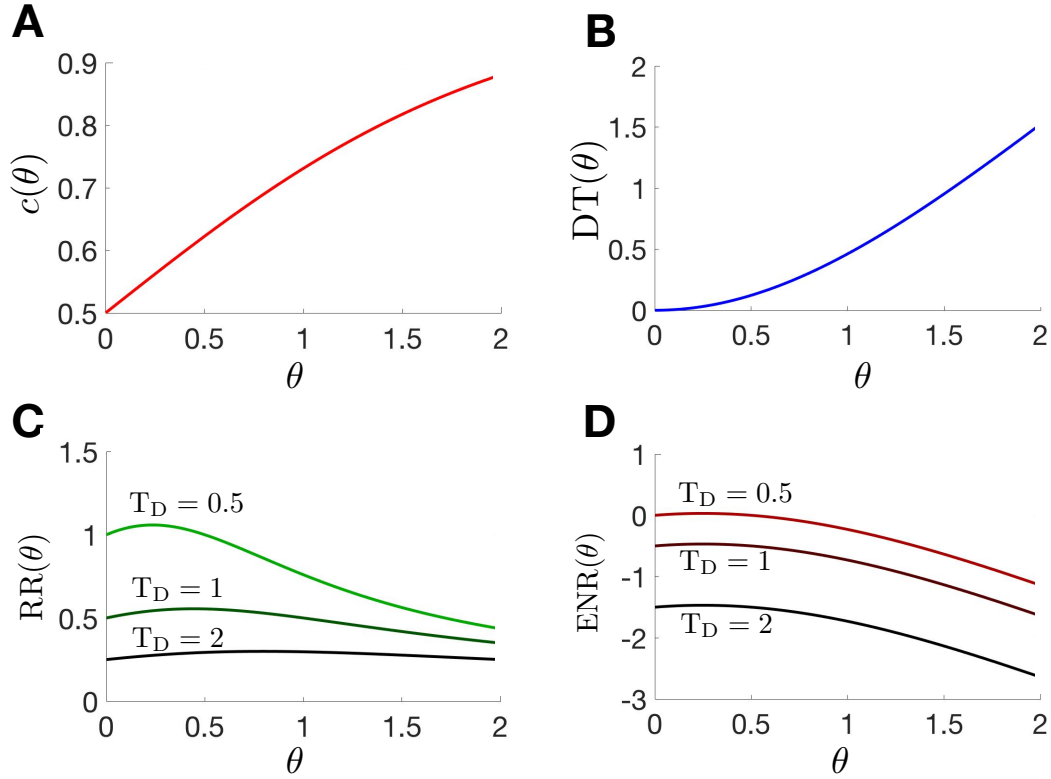


Fig. 2.4.1: The variance is $D = 1$. **A.** Correct probability and **B.** decision time as functions of decision thresholds θ . As θ increases, it takes more time for the observer to make decisions and the correct probability is improved. **C.** Reward rate (RR) and **D.** expected net reward (ENR) as functions of thresholds θ , where the “cost” = 1. The delay time T_D has an impact on the optimum of RR. In contrast, because of the linearity, T_D only causes a vertical shift in ENR.

2.4.2 Optimizing Expected Net Reward

Another quantity that can reflect the speed-accuracy tradeoff is the expected net reward (ENR) function [23]

$$\text{ENR}(\theta) = c - \text{cost} \cdot (\text{DT} + T_D), \quad (2.4.3)$$

where c , DT , and T_D are defined in section 2.4.1. The “cost” variable is the cost of the evidence accumulation process or energy that an observer spends collecting information. In fact, if the observer takes too long to decide on one trial, they will have less time for future trials. The higher the cost, the fewer rewards will be given. The cost of evidence accumulation plays an important role because without it, the observer will want to wait infinitely to get the most reward. The cost could be implicit due to one’s effort or attention or explicit such as losing future reward [23]. In contrast to the optimal threshold of the RR, the ENR optimal threshold does not depend on the intertrial delay time T_D . Changing the delay time merely shifts the ENR up and down (See Fig. 2.4.1 D).

2.5 Summary

We have reviewed the derivation of the standard DDM from first principles and computed its statistics. In addition, we have shown how to optimize the observer’s performance in a single trial using two different measurements: (1) reward rate, and (2) expected net reward. Even though the standard DDM explains the decision-making process within a single trial well, it fails to take correlation

2.5. SUMMARY

and feedback into account. In nature, both animals and humans exhibit correlations in their trial-to-trial response patterns. Feedback from the natural world helps them learn the latent structure of their environment. In the next chapters, we develop a more realistic normative model for decision-making by considering intertrial correlations and feedback.

Integrating Information Across Two Correlated Trials

We model a repeated 2AFC 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 $P(H^1 = H_{\pm}) = 0.5$ and $P(H^j = H_{\mp} | H^{j-1} = H_{\pm}) = \epsilon$ for $j = 2, \dots, n$ (See Fig. 3.0.1). When $\epsilon = 0.5$ the states of subsequent trials are independent, and the decision on one trial does not bias future decisions for an ideal observer [8, 38, 39, 71, 73, 83].

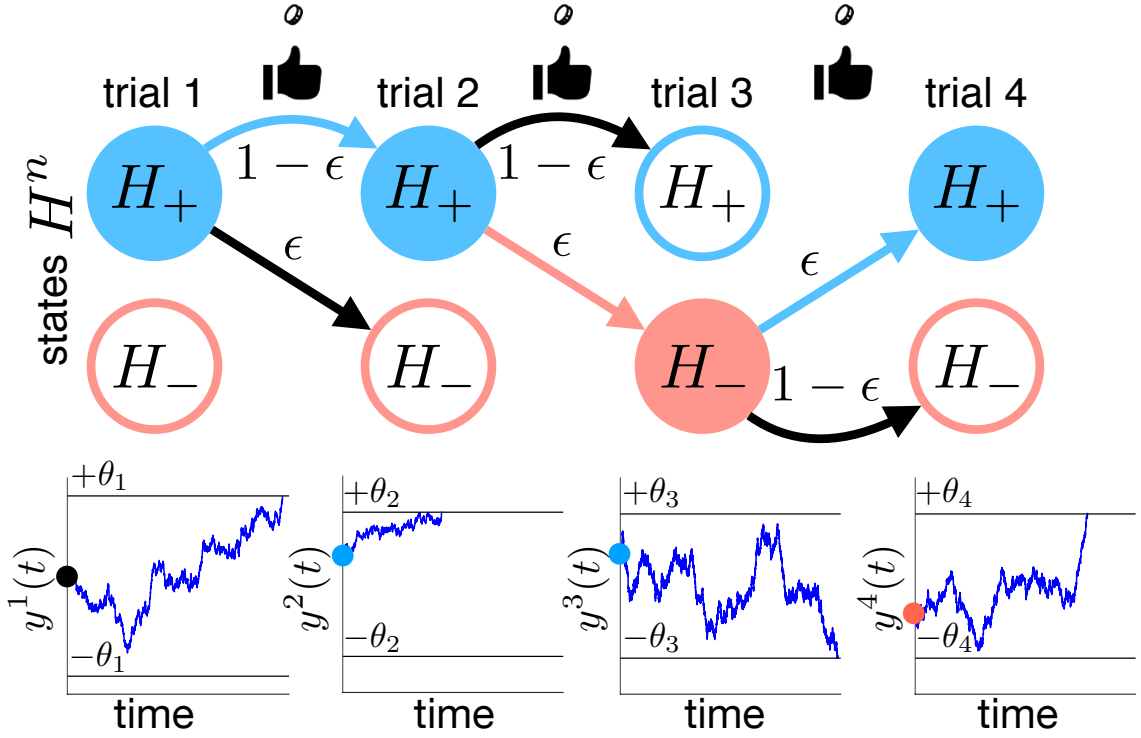


Fig. 3.0.1: Sequence of 2AFC trials with states H^n determined by a two-state Markov process with switching probability $\epsilon := P(H^n \neq H^{n-1})$. For instance, $H^{1:4} = (H_+, H_+, H_-, H_+)$ stands for the true states of the environment. When the environment is H_+ (H_-), the drift direction of the drift-diffusion process is positive (negative). The decision threshold (θ_{n-1}) crossed on trial $n - 1$ determines the sign of the initial condition, $y^n(0)$, on trial n .

When repetitions are more likely and trials are dependent, i.e., $0 \leq \epsilon < 0.5$, evidence obtained during one trial can inform a decision on the next trial. If $0.5 < \epsilon \leq 1$, states are more likely to alternate. The analysis is similar to the one we present here, so we do not discuss this case separately. We will show that ideal observers use their decision on the previous trial to adjust their 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 [11]. Rewarding the correct choice on trial j provides unambiguous information about the state H^j , superseding the noisy information gathered over that trial. The results can be easily extended to this case, as well as cases when rewards are only given with some probability, as we show in chapter 5.

Our main contribution is to derive and analyze the ideal observer model for this sequential version of the 2AFC 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.

We first focus on the case of two sequential 2AFC trials. Both states are equally likely on the first trial, $P(H^1 = H_{\pm}) = 0.5$, but the state on the second trial can depend on the state of the first, $\epsilon := P(H^2 = H_{\mp} | H^1 = H_{\pm})$. On trial 2, the ideal observer makes observations, $\xi_{1:m_2}^2$, with conditional density $f_{\pm}(\xi_{1:m_2}^2)$ if $H^2 = H_{\pm}$ to infer the state H^2 . All information about H^1 can be obtained from the decision variable, $d_1 = \pm 1$, and the ideal observer uses 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.

3.1 Optimal Observer's Priors

The first trial is equivalent to the single trial case discussed in chapter 2. Assuming each choice is equally likely, $P(H^1 = H_{\pm}) = 1/2$, no prior information exists. Therefore, the decision variable, $y^1(t)$, evolves according to the DDM given by Eq. (2.0.1) 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. (2.3.5). Furthermore, since $\epsilon = P(H^2 = H_{\mp} | H^1 = H_{\pm})$, it follows that $1 - \epsilon = P(H^2 = H_{\pm} | H^1 = H_{\pm})$, which means

$$\begin{aligned} P(H^2 = H_{\pm} | d_1 = \pm 1) &= P(H^2 = H_{\pm} | H^1 = H_{\pm}) P(H^1 = H_{\pm} | d_1 = \pm 1) \\ &\quad + P(H^2 = H_{\pm} | H^1 = H_{\mp}) P(H^1 = H_{\mp} | d_1 = \pm 1) \\ &= (1 - \epsilon)c_1 + \epsilon(1 - c_1), \end{aligned} \tag{3.1.1}$$

and, similarly,

$$P(H^2 = H_{\mp} | d_1 = \pm 1) = (1 - \epsilon)(1 - c_1) + \epsilon c_1. \tag{3.1.2}$$

As the decision, $d_1 = \pm 1$, determines the probability of each state at the end of trial 1, individual observations, $\xi_{1:m_1}^1$, 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, $\xi_{1:m_2}^2$, and their decision on the previous trial, d_1 , to arrive at the

probability ratio

$$R_{m_2}^2 = \frac{P(H^2 = H_+ | \xi_{1:m_2}^2, d_1)}{P(H^2 = H_- | \xi_{1:m_2}^2, d_1)} = \frac{P(\xi_{1:m_2}^2 | H^2 = H_+) P(H^2 = H_+ | d_1)}{P(\xi_{1:m_2}^2 | H^2 = H_-) P(H^2 = H_- | d_1)}.$$

Taking the logarithm, and applying conditional independence of the measurements $\xi_{1:m_2}^2$, we have that

$$L_{m_2}^2 = \ln R_{m_2}^2 = \sum_{j=1}^{m_2} \ln \frac{f_+(\xi_j^2)}{f_-(\xi_j^2)} + \ln \frac{P(H^2 = H_+ | d_1)}{P(H^2 = H_- | d_1)},$$

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 2.1, we find

$$dy^2 = g^2 dt + \sqrt{2D} dW, \quad (3.1.3)$$

with the Wiener process, W , and variance defined as in Eq. (2.0.1). 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 are correlated across trials. The initial condition for Eq. (3.1.3) can be derived from Eq. (2.1.4) and is therefore

$$y^2(0; d_1 = \pm 1) = D \ln \frac{P(H^2 = H_+ | d_1 = \pm 1)}{P(H^2 = H_- | d_1 = \pm 1)} = \pm D \ln \frac{(1 - \epsilon)e^{\theta_1/D} + \epsilon}{\epsilon e^{\theta_1/D} + (1 - \epsilon)} = \pm y_0^2, \quad (3.1.4)$$

where we have used Eqs. (2.3.5), (3.1.1), and (3.1.2). Note that $y_0^2 \rightarrow 0$ in the limit $\epsilon \rightarrow 0.5$, so no information is carried forward when uncorrelated across trials. In

3.2. PERFORMANCE ON THE FIRST TWO TRIALS

the limit $\epsilon \rightarrow 0$, we find $y_0^2 \rightarrow \theta_1$, so the ending value of the decision variable $y^1(T_1) = \pm\theta_1$ on 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 $y_0^2 \in (0, \theta_1)$ [61].

We assume that the decision variable, $y^2(t)$, evolves according to Eq. (3.1.3) 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

$$\begin{aligned} dy^j &= g^j dt + \sqrt{2D} dW, \quad \text{until } |y^j(T_j)| \geq \theta_j \mapsto d_j = \text{sign}(y^j(T_j)), \\ y^1(0) &= 0, \quad y^2(0) = \pm y_0^2 \quad \text{for } d_1 = \pm 1, \quad \text{and} \\ g^j &= \pm 1 \quad \text{for } H^j = H_{\pm}, j = 1, 2. \end{aligned}$$

3.2 Performance on the First Two Trials

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

$$\text{RR}_{1:2} = \frac{c_1 + c_2}{\text{DT}_1 + \text{DT}_2 + 2\text{T}_D}, \quad (3.2.1)$$

where c_j is the probability of correct responses on trial $j = 1, 2$, DT_j is the mean decision time on 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 on trial 2 will depend on the outcome of trial 1.

3.2.1 Threshold Determines the Probability of Being Correct

We note that Eq. (2.1.4) 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

$$\frac{P(H^2 = H_+ | y^2(T_2) = \theta)}{P(H^2 = H_- | y^2(T_2) = \theta)} = e^{\theta/D}, \quad (3.2.2)$$

so that rearranging Eq. (3.2.2) yields

$$c_2 := \frac{P(H^2 = H_+ | y^2(T_2) = \theta)P(y^2(T_2) = \theta)}{P(d_2 = +1)} = \frac{1}{1 + e^{-\theta/D}} = c_1. \quad (3.2.3)$$

Therefore the probability of correct decision 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 . On trial 2, information from the previous trial is already incorporated as a prior, so the threshold directly determines the probability of a correct decision on both trials 1 and 2.

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. (3.2.3) begins by computing the total probability of a correct choice on trial 2 by combining the conditional probabilities

3.2. PERFORMANCE ON THE FIRST TWO TRIALS

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 $-y_0^2$

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

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

3.2.2 Mean Response Time on the Second Trial

We can compute the expected time until a decision on trial 2, DT_2 , by marginalizing over $d_1 = \pm 1$,

$$DT_2|(H^2 = H_{\pm}) = \sum_{k=\pm 1} T(k \cdot y_0^2; \theta_2) \cdot P(d_1 = k|H^2 = H_{\pm}), \quad (3.2.4)$$

3.2. PERFORMANCE ON THE FIRST TWO TRIALS

since $d_1 = \pm 1$ will lead to the initial condition $\pm y_0^2$ on trial 2. We can compute $T(\pm y_0^2; \theta)$ using Eq. (2.2.3) as

$$T(\pm y_0^2; \theta_2) = \theta_2 \left[\coth \frac{\theta_2}{D} - \operatorname{csch} \frac{\theta_2}{D} \left[\frac{\epsilon e^{\theta_1/D} + (1 - \epsilon)}{(1 - \epsilon) e^{\theta_1/D} + \epsilon} \right]^{\pm 1} \right] \mp D \ln \frac{(1 - \epsilon) e^{\theta_1/D} + \epsilon}{\epsilon e^{\theta_1/D} + (1 - \epsilon)}. \quad (3.2.5)$$

The second terms in the summands of Eq. (3.2.4) are given by

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

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

The average time until a decision on trial 2 is (see section 3.2.2)

$$DT_2 = \frac{1 - e^{-\theta/D}}{1 + e^{-\theta/D}} \left[\theta - (1 - 2\epsilon) D \ln \frac{(1 - \epsilon) e^{\theta/D} + \epsilon}{\epsilon e^{\theta/D} + (1 - \epsilon)} \right]. \quad (3.2.6)$$

Notice, in the limit $\epsilon \rightarrow 0.5$, Eq. (3.2.6) reduces to Eq. (2.2.4) for $\theta_1 = \theta$ as expected. Furthermore, in the limit $\epsilon \rightarrow 0$, $DT_2 \rightarrow 0$, since decision is 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. (2.3.5), and (3.2.6) with

Eq. (3.2.1) to find

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

which can be maximized using numerical optimization.

Our analysis shows that the correct probability on both trials, $c_{1:2}$, increases with θ and does not depend on the transition probability ϵ (Fig. 3.2.1A). In addition, the decision time on trial 2, DT_2 , increases with θ and ϵ (Fig. 3.2.1B): At smaller transition rates, ϵ , more information is carried forward to the second trial, so the decision variable y_0^2 is, on average, closer to the threshold θ . This shortens the average decision time and increases the RR (Fig. 3.2.1C, 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.

3.2. PERFORMANCE ON THE FIRST TWO TRIALS

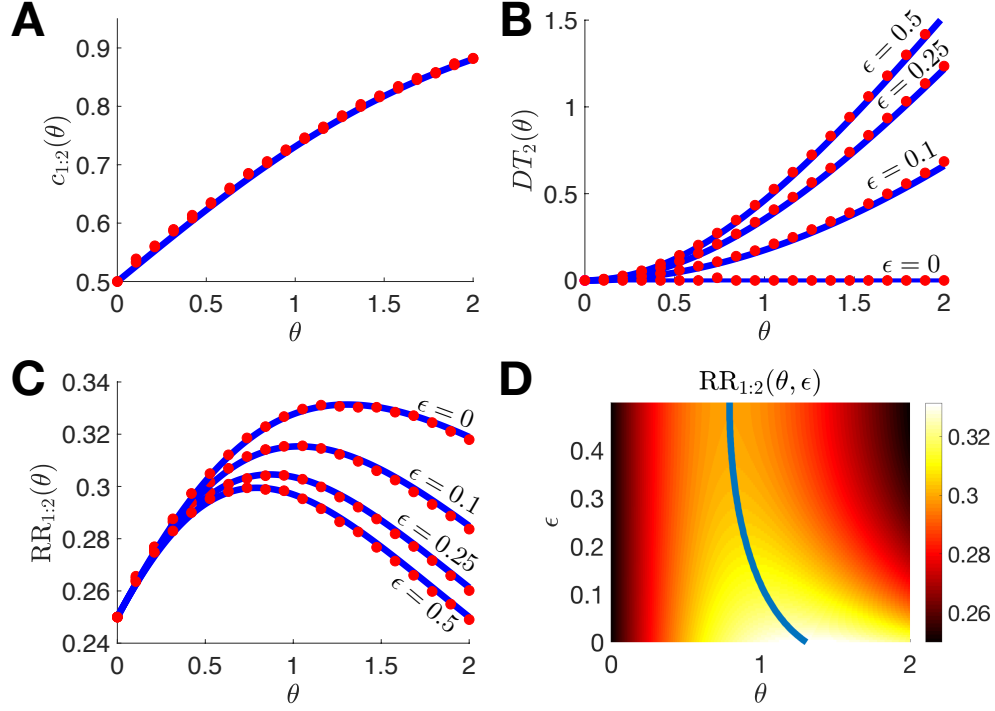


Fig. 3.2.1: The reward rate (RR) depends on the state change rate $\epsilon := P(H^2 \neq H^1)$ for a constant threshold $\theta := \theta_{1:2}$. Here and below stochastic simulations (dots) are compared to theoretical predictions (solid lines). **A.** The probability of a correct response on trials 1 and 2 ($c_{1:2}$) only depends on the threshold θ , and increases with θ . **B.** Decision time DT_2 on 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. (3.2.1) 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 time on trial 1 because 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 $D = 1$.

3.3 Performance for Dynamic Decision Thresholds

We next ask whether the RR given by Eq. (3.2.1) 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 , on trial 1 are given by Eqs. (2.2.4) and (2.3.4). The correct probability c_2 on 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 on 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

$$c_2 = \begin{cases} \frac{1}{1 + e^{-\theta_2/D}} & : \theta_2 > y_0^2, \\ \frac{(1 - \epsilon) + \epsilon e^{-\theta_1/D}}{1 + e^{-\theta_1/D}} & : \theta_2 \leq y_0^2. \end{cases} \quad (3.3.1)$$

In the case $\theta_2 > y_0^2$, the average time until a decision on trial 2, is (See 3.2.2)

$$DT_2 = \theta_2 \frac{1 - e^{-\theta_2/D}}{1 + e^{-\theta_2/D}} - \frac{(1 - 2\epsilon)(1 - e^{-\theta_1/D})}{1 + e^{-\theta_1/D}} D \ln \frac{(1 - \epsilon)e^{\theta_1/D} + \epsilon}{\epsilon e^{\theta_1/D} + (1 - \epsilon)}, \quad (3.3.2)$$

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

$$RR_{1:2}(\epsilon) = \begin{cases} \frac{E_+^1 + E_+^2}{E_+^2 E_-^1 [\theta_1 - (1 - 2\epsilon)y_0^2] + E_+^1 [\theta_2 E_-^2 + 2T_D E_+^2]} & : \theta_2 > y_0^2, \\ \frac{2(1 - \epsilon) + \epsilon E_+^1}{\theta_1 E_-^1 + 2T_D E_+^1} & : \theta_2 \leq y_0^2, \end{cases} \quad (3.3.3)$$

3.3. PERFORMANCE FOR DYNAMIC DECISION THRESHOLDS

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.

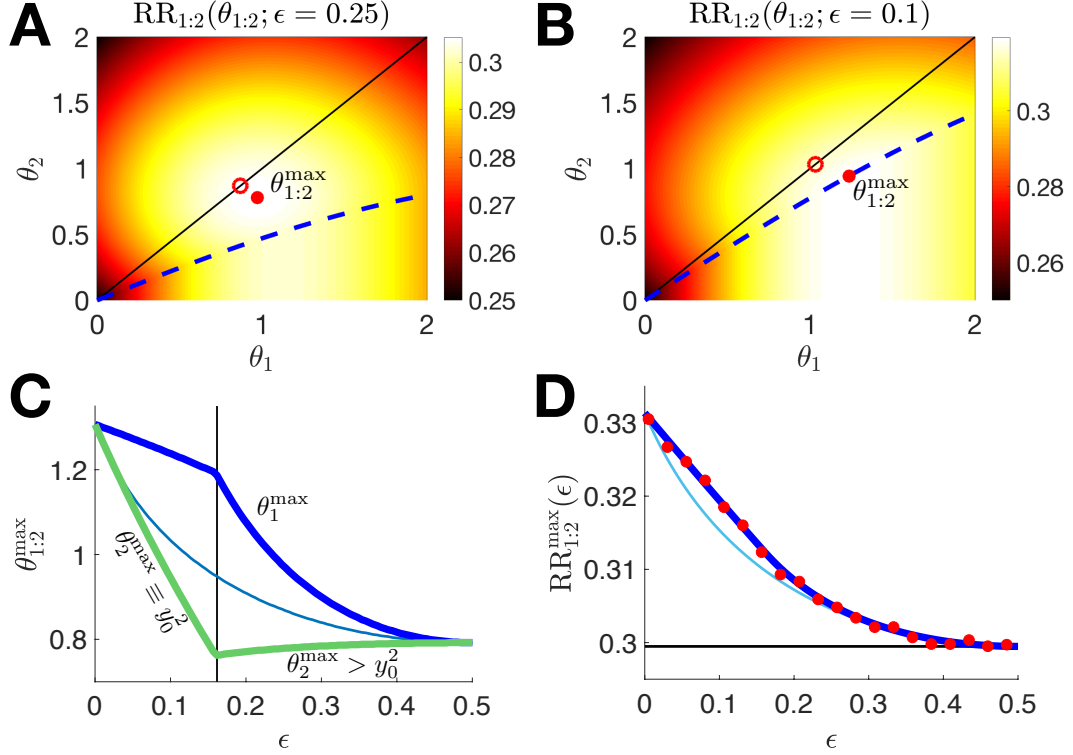


Fig. 3.3.1: **A,B.** Reward rate (RR), given by Eq. (3.3.3), depends on the change rate ϵ and thresholds $\theta_{1:2}$. For $\epsilon = 0.25$ (panel **A**), the pair $\theta_{1:2}^{\max}$ that maximizes RR (dot) satisfies $\theta_1^{\max} > \theta_2^{\max}$. Compare to the pair $\theta_{1:2}^{\max}$ (circle) for $\theta = \theta_{1:2}$ as in Fig. 3.2.1D. The boundary $\theta_2 = y_0^2(\theta_1)$, obtained from Eq. (3.1.4), is shown as a dashed line. For $\epsilon = 0.1$ (panel **B**), the optimal choice of thresholds (dot) is on the boundary $\theta_2 = y_0^2(\theta_1)$. **C.** Decreasing ϵ from 0.5, initially results in an increase in $\theta_1^{\max}$ and decrease in $\theta_2^{\max}$. At sufficiently small value of ϵ , the second threshold collides with the initial condition, $\theta_2 = y_0^2(\theta_1)$ (vertical line). After this point, both $\theta_1^{\max}$ and $\theta_2^{\max}$ increase with ϵ and meet at $\epsilon = 0$. The optimal constant threshold, $\theta^{\max}$, is represented by the thin curve. **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 light curve shows $RR_{1:2}^{\max}$ for the constant threshold case, and the horizontal line shows $RR_{1:2}^{\max}$ when assuming independence.

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. 3.3.1A,B). This shortens the mean time to a decision on 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 $\theta_2^{\max}$ decreases and $\theta_1^{\max}$ increases (Fig. 3.3.1C), and $\theta_2^{\max}$ eventually collides 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. 3.3.1D), 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. 3.2.1 and 3.3.1), 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 is slightly more pronounced when T_D is small or when D is large.

3.4 Optimizing the Expected Net Reward

Our next question is what is the optimal strategy if the performance is defined as ENR. From Eq. (3.3.2) and (3.3.1), we can write ENR function as the following

3.4. OPTIMIZING THE EXPECTED NET REWARD

$$\text{ENR}_{1:2} = \begin{cases} \frac{E_+^1 + E_+^2 - \text{cost} \cdot [E_+^2 E_-^1 (\theta_1 - \gamma(1-2\epsilon)y_0^2) + E_+^1 [\gamma\theta_2 E_-^2 + (1+\gamma)T_D E_+^2]]}{E_+^1 E_+^2} & : \theta_2 > y_0^2, \\ \frac{1 + \gamma(1 - \epsilon E_-^1) - \text{cost} \cdot [\theta_1 E_-^1 + (1+\gamma)T_D E_+^1]}{E_+^1} & : \theta_2 \leq y_0^2, \end{cases} \quad (3.4.1)$$

where $E_{\pm}^j = 1 \pm e^{-\theta_j/D}$, and $0 \leq \gamma \leq 1$ is the discount rate. The discount rate reflects the current value of future rewards: the reward is received j trials later in the future is only worth γ^{j-1} times of its immediate value [87].

We analyze ENR similarly to the way we do RR analysis. First, we look at the color map of ENR as a function of θ_1 and θ_2 for 2 different values of change rate, $\epsilon = 0.1$ and $\epsilon = 0.25$. In both cases, the global maximum of ENR is obtained when threshold on trial 2 is lower, i.e., $\theta_2 < \theta_1$ (see Fig. 3.4.1 A, B). This is the same result as we find in section 3.3 (see Fig. 3.3.1 A and B). Interestingly, we also find similar trends when we plot optimal thresholds $\theta_{1:2}^{\max}$ and maximum ENR as a function of ϵ (compare Fig. 3.4.1 A and B to Fig. 3.3.1 A and B, respectively). The same explanation for RR can be explained here. When initial belief collides with the threshold on trial 2, the observer makes an immediate decision. Therefore, increasing θ_2 does not impact the decision time while improving the correct probability. We also find that the dynamic threshold strategy improves slightly for small values of change rate.

3.4. OPTIMIZING THE EXPECTED NET REWARD

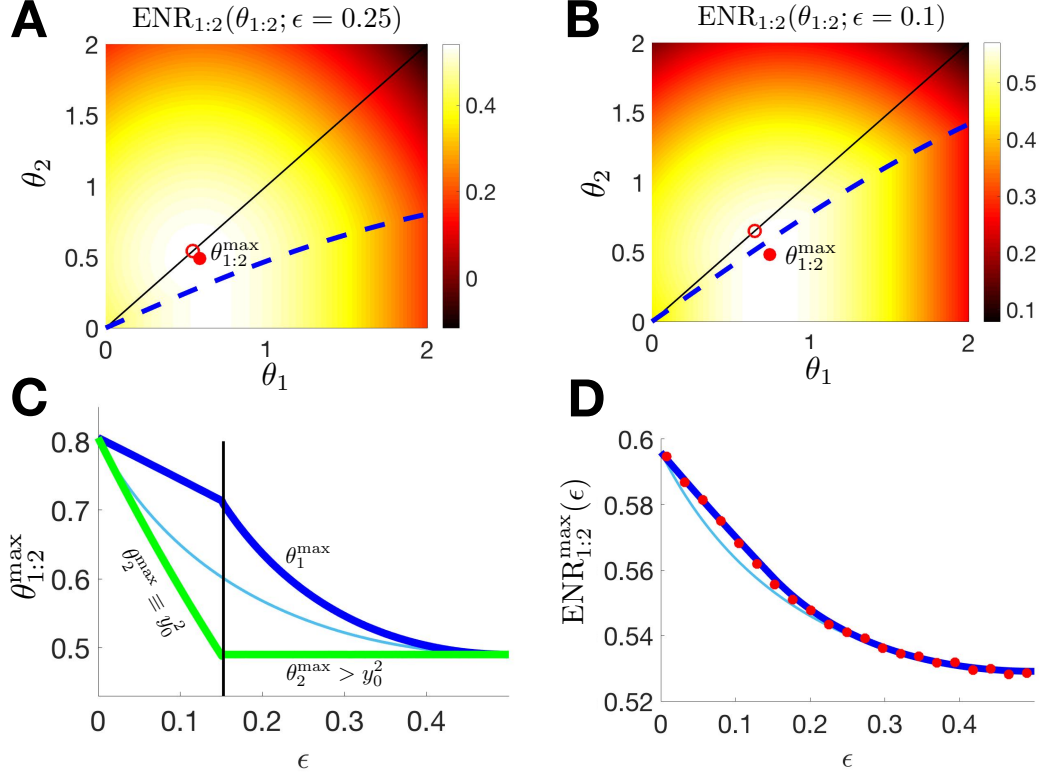


Fig. 3.4.1: Diffusion term $D = 1$, the evidence-accumulation cost is “cost” = 0.5, the intertrial delay time is $T_D = 0.5$, and discount rate $\gamma = 0.7$. **A, B.** Expected Net Reward (ENR), given by Eq. (3.4.1) as a function of change rate ϵ and thresholds $\theta_{1:2}$. The pair $\theta_{1:2}^{\max}$ that maximizes ENR (dot) satisfies $\theta_1^{\max} > \theta_2^{\max}$. On the other hand, the pair $\theta_{1:2}^{\max}$ (circle) maximizes ENR under the constraint on constant thresholds, $\theta_1 = \theta_2$. The boundary $\theta_2 = y_0^2(\theta_1)$, obtained from Eq. (3.1.4), is shown as a dashed line. **C, D.** Optimal decision thresholds $\theta_{1:2}^{\max}$ and maximum $\text{ENR}_{1:2}^{\max}$ as functions of change rate ϵ . The trends are similar to the results in Fig. 3.3.1 **C** and **D** even though the optimal values are different between RR and ENR. The thin light blue curve shows $\text{ENR}_{1:2}^{\max}$ for the constant threshold case, and the vertical line in panel **C** shows the boundary of $\theta_2 = y_0^2(\theta_1)$.

Remark: The optimal thresholds found from optimizing ENR are lower than from RR optimization. Recall that the intertrial delay time T_D makes a difference in optimizing RR while it does not affect optimizing ENR. We set high delay time

($T_D = 2$) in our RR optimization problem, so it increases the optimal thresholds. Note that ENR and RR are two different functions, so we cannot say that one function is a better optimization.

3.5 Summary

In this chapter, we extend the standard DDM to include correlation while omitting feedback. We compute the history bias, mean decision time, and probability of being correct. Optimizing RR and ENR for the first two trials, we find that varying thresholds across trials yields the global maximum. In fact, the optimal strategy is to set higher thresholds on trial 1 and lower thresholds on trial 2 so that the bias is closer to the threshold. In chapter 4, we extend these observations to multiple correlated trials.

Integrating Information Across 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 $\epsilon := P(H^{j+1} = H_{\mp} | H^j = H_{\pm})$. 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, $\theta_j^{\max} \leq \theta_{j-1}^{\max}$.

4.1 Optimal Observer Model for Multiple Trials

As in the case of two trials, we must consider the caveats associated with instantaneous decisions 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. (3.1.3) shows that the decision variable evolves according to a linear drift-diffusion process on each trial. Because the thresholds are symmetric, we only discuss the case where the correct threshold is positive. The probability of a correct choice is again defined by the threshold, $c_j = 1/(1 + e^{-\theta_j/D})$ when $\theta_j > y_0^j$. When $\theta_j \leq y_0^j$, no evidence is accumulated and a decision is made immediately. The correct probability is $c_j = (1 - \epsilon)c_{j-1} + \epsilon(1 - c_{j-1})$ as in Eq. (3.3.1). Therefore, c_j may be defined iteratively for $j \geq 2$, as

$$c_j = \begin{cases} \frac{1}{1 + e^{-\theta_j/D}} & : \theta_j > y_0^j, \\ (1 - \epsilon)c_{j-1} + \epsilon(1 - c_{j-1}) & : \theta_j \leq y_0^j, \end{cases} \quad (4.1.1)$$

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

The probability c_j also quantifies the belief of a rational observer at the end of trial j . Hence, as in Eq. (3.1.4), the initial belief of an ideal observer on trial $j + 1$ has the form

$$y^{j+1}(0; d_j = \pm 1) = \pm D \ln \frac{(1 - \epsilon)c_j + \epsilon(1 - c_j)}{\epsilon c_j + (1 - \epsilon)(1 - c_j)} = \pm y_0^{j+1}, \quad (4.1.2)$$

and the decision variable $y^j(t)$ within trials obeys a DDM with threshold θ_j as in Eq. (3.1.3). 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 $T_j = 0$. In the following we assume that the number of trials, n , is known to the observer.

4.2 Performance for Constant Decision Thresholds

We first assume that the observer uses a fixed threshold, $\theta_j = \theta$, across all trials. With n trials, the RR is the sum of the expected rewards divided by the total expected time for the sequence of trials. Assuming, as before, that correct responses yield a unit reward, and incorrect responses yield no reward, and that the total time is a sum of the decision and delay time, we obtain

$$\text{RR}_{1:n} = \frac{\sum_{j=1}^n c_j}{\sum_{j=1}^n \text{DT}_j + nT_D}. \quad (4.2.1)$$

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. (4.1.2) 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 trial 2 and beyond, the analysis of the two-trial case implies that the mean decision time is given by an expression equivalent to Eq. (3.2.6) for $j \geq 2$. Using these expressions for c_j and DT_j in Eq. (4.2.1), we find

$$\text{RR}_{1:n}(\theta; \epsilon) = n \left(\left[1 - e^{-\theta/D} \right] \left[n\theta - (n-1)(1-2\epsilon)y_0 \right] + nT_D \left[1 + e^{-\theta/D} \right] \right)^{-1}. \quad (4.2.2)$$

As n is increased, the threshold value, $\theta^{\max}$, that maximizes the RR increases

4.2. PERFORMANCE FOR CONSTANT DECISION THRESHOLDS

across the full range of ϵ values (Fig. 4.2.1) as 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) on later trials. In the limit of an infinite number of trials ($n \rightarrow \infty$), Eq. (4.2.2) simplifies to

$$\text{RR}_\infty(\theta; \epsilon) = \left(\left[1 - e^{-\theta/D} \right] [\theta - (1 - 2\epsilon)y_0] + T_D \left[1 + e^{-\theta/D} \right] \right)^{-1}, \quad (4.2.3)$$

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

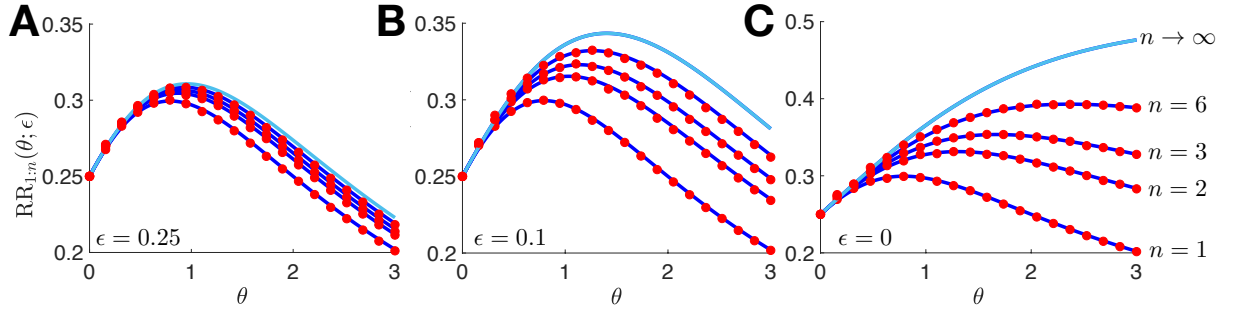


Fig. 4.2.1: Reward rate ($\text{RR}_{1:n}$) defined by Eq. (4.2.2) increases with n , and as ϵ is decreased, as shown for $\epsilon = 0.25$ (panel A); $\epsilon = 0.1$ (panel B); and $\epsilon = 0$ (panel C). The optimal decision threshold, $\theta^{\max}$, increases with n . The limiting reward rates $\text{RR}_{1:n}(\theta; \epsilon) \rightarrow \text{RR}_\infty(\theta; \epsilon)$ as $n \rightarrow \infty$ defined by Eq. (4.2.3) are given by the topmost light curves.

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

$$\lim_{\epsilon \rightarrow 0} \text{RR}_\infty(\theta; \epsilon) = \frac{1}{T_D [1 + e^{-\theta/D}]},$$

so the decision threshold value that maximizes RR diverges, $\theta^{\max} \rightarrow \infty$ (Fig. 4.2.1C). Intuitively, when there are many trials ($n \gg 1$), the price of a long wait for a high accuracy decision on the first trial ($c_1 \approx 1$ but $\text{DT}_1 \gg 1$ 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 $DT_j = 0$ for $\theta \gg 1$ and $j > 1$).

4.3 Performance for Dynamic Decision Thresholds

In the case of dynamic thresholds, the RR function, Eq. (4.2.1), 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. (4.1.1). Therefore, while analytical results can be obtained for constant thresholds, numerical methods are needed to find the sequence of optimal dynamic thresholds.

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

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

Thus, Eq. (4.1.1), 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. (4.2.1) to determine the RR which achieves a maximum for some sequence of thresholds, $\theta_{1:n} := (\theta_1, \theta_2, \dots, \theta_n)$.

In Fig. 4.3.1 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 of earlier trials improves accuracy on later trials. Thus, there is a value in gathering more information early in the trial sequence, and making fast, but less accurate decisions later in the 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, $\theta_n^{\max}$, 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 is made instantaneously. This results in a more gradual decrease in the thresholds used on each successive trial in the case of slow environments, with low ϵ values.

4.4. PERFORMANCE FOR DYNAMIC DECISION THRESHOLDS USING EXPECTED NET REWARD

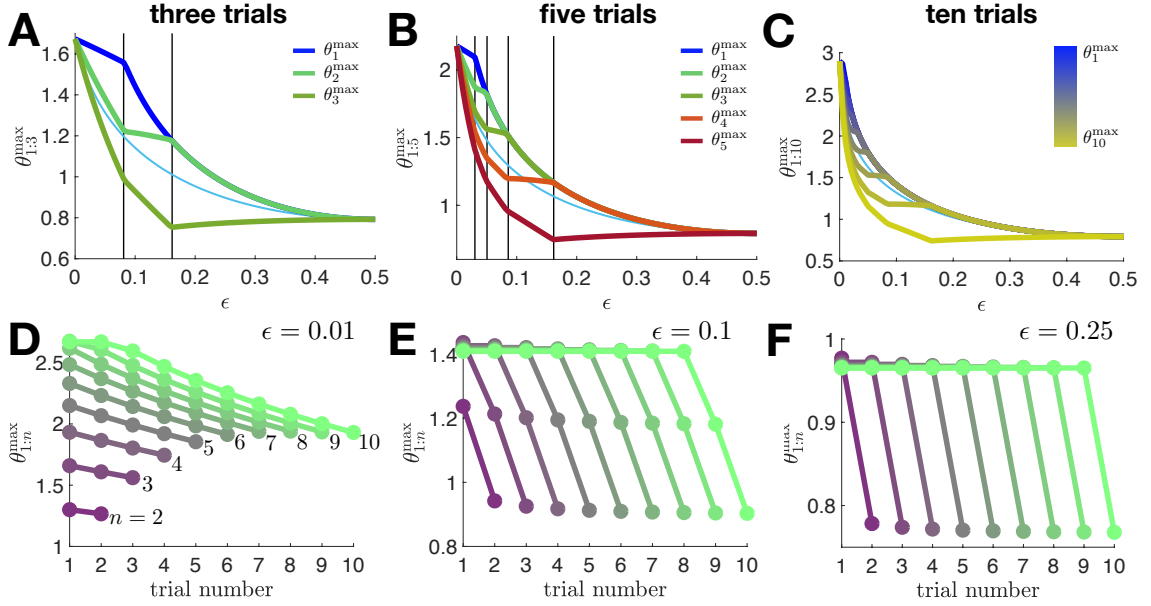


Fig. 4.3.1: Optimal sets of thresholds, $\theta_{1:n}^{\max}$, for n trials with correlated states. Top row: **A**. For $n = 3$, the set of optimal thresholds $\theta_{1:3}^{\max}$ obeys the ordering $\theta_1^{\max} \geq \theta_2^{\max} \geq \theta_3^{\max}$ for all values of ϵ , converging as $\epsilon \rightarrow 0.5^-$ 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**). Thin curves in each plot show the optimal choice of constant threshold θ . Bottom row: Fixing **D**. $\epsilon = 0.01$, **E**. $\epsilon = 0.1$, and **F**. $\epsilon = 0.25$, the optimal threshold vectors $\theta_{1:n}^{\max}$ are displayed as a function of trial number. The stereotyped decrease in $\theta_j^{\max}$ is preserved as a function of j across different total trial numbers n . The decrease is more gradual for small ϵ , whereas for larger ϵ , thresholds are constant until the last trial $\theta_n^{\max}$.

4.4 Performance for Dynamic Decision Thresholds Using Expected Net Reward

We now identify optimal evidence accumulation strategies across trials when taking the expected net reward (ENR) as the value function, which for n trials is

4.4. PERFORMANCE FOR DYNAMIC DECISION THRESHOLDS USING EXPECTED NET REWARD

defined as

$$\text{ENR}_{1:n} = \sum_{j=1}^n \gamma^{j-1} [c_j - \text{cost} \cdot (\text{DT}_j + T_D)] \quad (4.4.1)$$

where c_j and DT_j are defined in Eq. (4.1.1) and (4.4.1). We assume the number of trials is finite, $n < \infty$, and the thresholds can vary across trials.

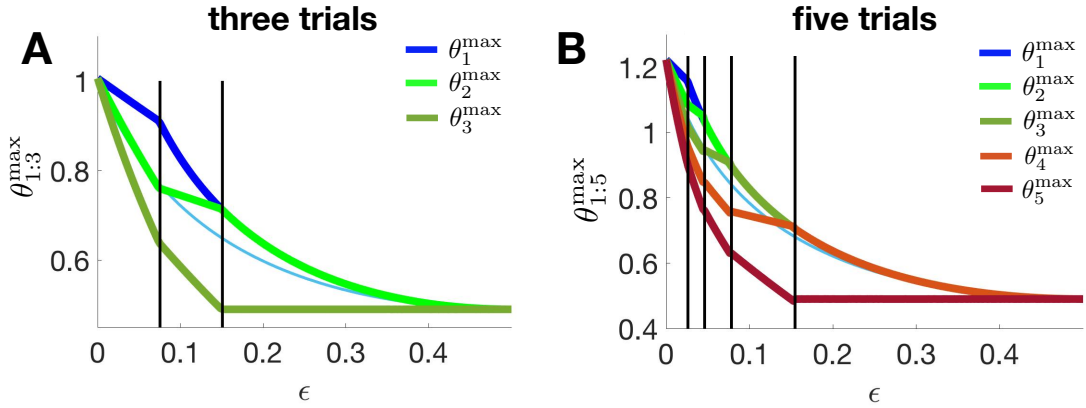


Fig. 4.4.1: Optimal sets of thresholds computed from ENR, $\theta_{1:n}^{\max}$, for n trials with correlated states. Here, $D = 1$, $\text{cost} = T_D = 0.5$ and $\gamma = 0.7$. **A.** For $n = 3$, the set of optimal thresholds $\theta_{1:3}^{\max}$ still obeys the ordering $\theta_1^{\max} \geq \theta_2^{\max} \geq \theta_3^{\max}$ 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.** For $n = 5$, the ordering of optimal thresholds is consistent, i.e., higher thresholds on earlier trials. Thin curves in each plot show the optimal choice of constant threshold θ .

When the observer optimizes ENR over multiple trials, we obtain the same result as when the observer optimizes RR (compare Fig. 4.3.1 A and B to Fig. 4.4.1 A and B, respectively): While the exact values of the thresholds are not the same in both cases, the thresholds decrease in the order from earlier trials to later trials, $\theta_1^{\max} \geq \theta_2^{\max} \geq \dots \geq \theta_n^{\max}$. This shows that the general optimal strategy is the

qualitatively regardless of whether the observer optimizes ENR or RR. As such, we conclude that both value functions quantify performance in such multi-trial tasks similarly enough that we need only focus on one henceforth. In the subsequent chapter, we will primarily focus on optimal strategy calculations using ENR as it is more tractable to analyze for some problems, due to the time penalty arising linearly, as opposed to nonlinearly in the RR.

4.5 Summary

In this chapter, we derive general equations for the probability of a correct response and the mean decision time for any number of trials. When the ideal observer changes their threshold between trials, their initial beliefs for a trial can be above thresholds, resulting in an immediate decision. We analyzed the no-feedback case in multiple trials by optimizing reward rate (RR) and expected net reward (ENR) functions. Even though both functions do not produce the exact same set of optimal thresholds, the trends in strategy are similar. The optimal observer prioritizes accuracy on earlier trials and speed on later trials. In the next chapter, we introduce probabilistic feedback and identify how conflict arises between external feedback and internal belief and how this shapes decision strategies.

Effects of Probabilistic Feedback Across Correlated Trials

Having studied the normative model of sequential decision-making without feedback, we next consider how the observer integrates external feedback with their internal belief to make optimal decisions in a changing environment. To model this situation we assume the following: the ideal observer integrates evidence about the true state, H^n , and makes a decision about whether $H^n = H_+$ or H_- . Feedback is then given in two possible ways. *Probabilistic signaling* denotes given feedback after the response as a light signal: green light implying the response was more likely correct; red light implying the response was more likely incorrect. As in the previous chapter, we assume that the reward is provided at the end of the trial sequence as it would provide an unambiguous signal that the response was correct. Alternatively, the reward itself may be provided at the end of each trial, but with some probability of being received when the observer is

wrong, or not received when the observer is right. We call this the case of *probabilistic reward* (See Fig. 5.0.1 for more details).

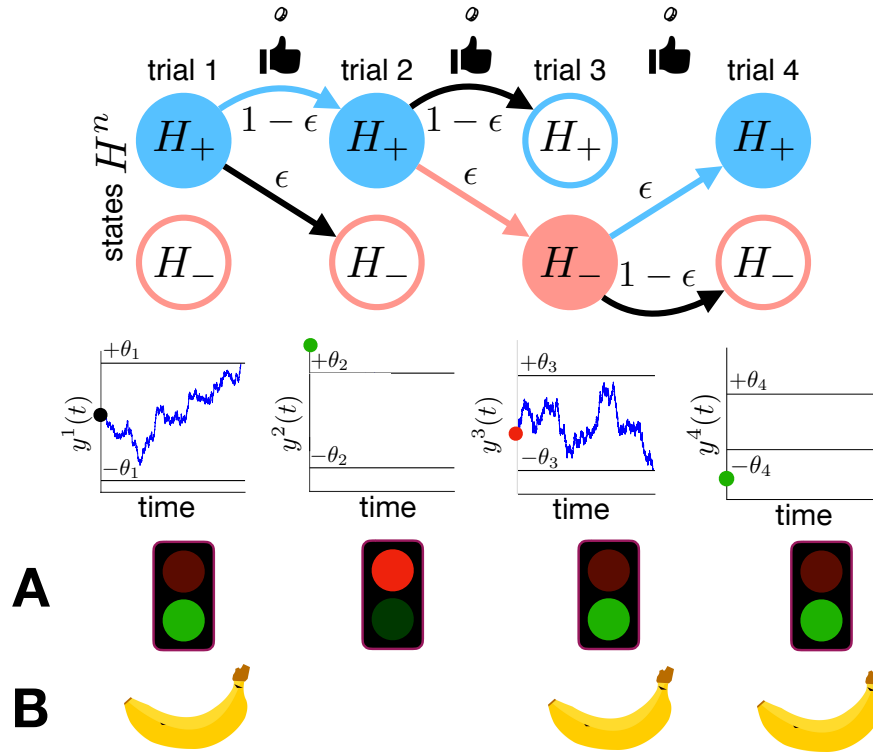


Fig. 5.0.1: True environmental state $H^n \in H_{\pm}$ switches stochastically between trials with probability ϵ . An observer sequentially accumulates noisy evidence within a trial and biases their initial belief based on the past choice and feedback. When the green light (or reward) appears at the end of trial 1, the initial belief on trial 2 may end up above the decision boundary due to the additional positive information. Thus, the observer may use feedback at the end of trial 1 to make an immediate decision on trial 2. **A.** Given probabilistic signal as feedback, despite the erroneous red (negative) signal at the end of trial 2, the observer will still receive a reward for trial at the end of the trial sequence, since they were correct. **B.** In the case of probabilistic reward, the observer is sometimes not rewarded/rewarded even though they are correct/incorrect. In this example, there is no reward on trial 2 even though the observer is correct.

Besides the change probability ϵ , we assume that the observer also knows the

probability of correct feedback, i.e., $p := P(\text{positive feedback}|\text{correct choice})$ and $q := P(\text{negative feedback}|\text{incorrect choice})$, i.e.,

$$P(s_j = +1|H^j = H_{\pm}, d_j = \pm 1) = p,$$

$$P(s_j = -1|H^j = H_{\pm}, d_j = \mp 1) = q,$$

where $s_j = +1$ ($s_j = -1$) indicates positive feedback (negative feedback) on j -th trial.

Rewards are distributed differently depending on the type of feedback. If the feedback is a probabilistic signal, a cumulative reward is distributed at the end of the sequence of trials. On the other hand, if the feedback is a reward, then reward is distributed probabilistically and does not perfectly correspond to the correctness or incorrectness of the observer's choices. For example, in the case of probabilistic signaling feedback, despite the red signal, a reward will still be given for the response on trial 2 at the end of all trials if the observer is correct (See Fig. 5.0.1 A). However, in the case of probabilistic reward, there may be no reward given at times even though an observer is correct on a trial (See Fig. 5.0.1 B). When $p = q = 0.5$, probabilistic signals are uninformative, but the observer can still be rewarded on more than half the trials on average by accumulating information, as described in chapters 3 and 4. On the other hand, for $p = q = 0.5$, probabilistic rewards are awarded independently of the responses. Observers are therefore rewarded only on half the trials on average, regardless of their strategy.

5.1 Feedback Enhances the Observer's Initial Belief

We start off by considering the first two trials of a sequence to get a sense of how information is passed from one trial to the next and how this impacts decision strategies. The observer combines their internal belief and external feedback on trial 1 to bias their belief on trial 2. We assume that the probabilities of correct feedback are the same for both positive and negative feedback, i.e., $p = q$, and decision thresholds are fixed across trials, $\theta_1 = \theta_2 = \theta$. We only consider the case where $0.5 \leq p \leq 1$. If $0 \leq p < 0.5$, the roles of positive feedback and negative feedback interchange, i.e., the red light (green light) indicates correct (incorrect) choice while the observer receives rewards for being incorrect more often. The analysis is similar, so we do not discuss this case separately.

Moreover, we have $\epsilon = P(H^2 = H_{\mp} | H^1 = H_{\pm}) = P(H^1 = H_{\pm} | H^2 = H_{\mp})$ and $1 - \epsilon = P(H^2 = H_{\pm} | H^1 = H_{\pm})$, then

$$\begin{aligned}
 & P(s_1 = +1, d_1 = \pm 1 | H^2 = H_{\pm}) \\
 &= P(s_1 = +1 | d_1 = \pm 1, H^1 = H_{\pm}) P(d_1 = \pm 1 | H^1 = H_{\pm}) (1 - \epsilon) \\
 &+ P(s_1 = +1 | d_1 = \pm 1, H^1 = H_{\mp}) P(d_1 = \pm 1 | H^1 = H_{\mp}) \epsilon \\
 &= p(1 - \epsilon)c_1 + (1 - p)\epsilon(1 - c_1).
 \end{aligned} \tag{5.1.1}$$

Similarly, we have

$$P(s_1 = +1, d_1 = \pm 1 | H^2 = H_{\mp}) = p\epsilon c_1 + (1 - p)(1 - \epsilon)(1 - c_1), \quad (5.1.2)$$

$$P(s_1 = -1, d_1 = \pm 1 | H^2 = H_{\pm}) = (1 - p)(1 - \epsilon)c_1 + p\epsilon(1 - c_1), \quad (5.1.3)$$

$$P(s_1 = -1, d_1 = \pm 1 | H^2 = H_{\mp}) = (1 - p)\epsilon c_1 + p(1 - \epsilon)(1 - c_1). \quad (5.1.4)$$

The probability ratio is:

$$\begin{aligned} R_{m_2}^2 &= \frac{P(H^2 = H_+ | \xi_{1:m_2}^2, d_1, s_1)}{P(H^2 = H_- | \xi_{1:m_2}^2, d_1, s_1)} \\ &= \frac{P(\xi_{1:m_2}^2 | H^2 = H_+) P(s_1, d_1 | H^2 = H_+)}{P(\xi_{1:m_2}^2 | H^2 = H_-) P(s_1, d_1 | H^2 = H_-)}, \end{aligned}$$

and thus, the LLR is computed as

$$L_{m_2}^2 = \ln R_{m_2}^2 = \sum_{j=1}^{m_2} \ln \frac{f_+(\xi_j^2)}{f_-(\xi_j^2)} + \ln \frac{P(s_1, d_1 | H^2 = H_+)}{P(s_1, d_1 | H^2 = H_-)},$$

indicating that $L_0^2 = \ln[P(s_1, d_1 | H^2 = H_+)/P(s_1, d_1 | H^2 = H_-)]$. Taking the continuum limit, we also find

$$dy^2 = g^2 dt + \sqrt{2D} dW. \quad (5.1.5)$$

When feedback is present, the initial belief is typically biased, but the direction of the bias depends on the strength of feedback p and discounting based on the

5.1. FEEDBACK ENHANCES THE OBSERVER'S INITIAL BELIEF

correct choice change probability ϵ . The initial condition for Eq. (5.1.5) is

$$\begin{aligned} y^2(0; d_1 = \pm 1, s_1 = +1) &= D \ln \frac{P(s_1 = +1, d_1 = \pm 1 | H^2 = H_+)}{P(s_1 = +1, d_1 = \pm 1 | H^2 = H_-)} \\ &= \pm D \ln \frac{(p - \epsilon)c_1 + (1 - p)\epsilon}{(p + \epsilon - 1)c_1 + (1 - p)(1 - \epsilon)} = \pm y_{0+}^2, \end{aligned} \quad (5.1.6)$$

$$\begin{aligned} y^2(0; d_1 = \pm 1, s_1 = -1) &= D \ln \frac{P(s_1 = -1, d_1 = \pm 1 | H^2 = H_+)}{P(s_1 = -1, d_1 = \pm 1 | H^2 = H_-)} \\ &= \pm D \ln \frac{(1 - p - \epsilon)c_1 + p\epsilon}{(\epsilon - p)c_1 + p(1 - \epsilon)} = \pm y_{0-}^2. \end{aligned} \quad (5.1.7)$$

When $p = q = 0.5$, the initial belief is

$$y^2(0; d_1 = \pm 1, s_1 = +1) = y^2(0; d_1 = \pm 1, s_1 = -1) = \pm D \ln \frac{(1 - \epsilon)c_1 + \epsilon(1 - c_1)}{\epsilon c_1 + (1 - \epsilon)(1 - c_1)},$$

which is the initial condition in the case of no feedback in Eq (4.1.2).

The observer enhances their initial belief when given probabilistic intertrial feedback. Reliable positive feedback, either probabilistic signaling or probabilistic reward, strengthens the observer's bias toward the same decision in a slow-changing environment, i.e., y_{0+}^2 increases toward $+\theta$ as $(\epsilon, p) \rightarrow (0, 1)$ (see Fig. 5.1.1 A and C). In contrast, reliable negative feedback strengthens the belief in opposite decision when the environmental state is likely to repeat, as y_{0-}^2 decreases toward $-\theta$ as $(\epsilon, p) \rightarrow (0, 1)$ (see Fig. 5.1.1 B and D). Interestingly, there is an increase in initial belief given by the red light when $(\epsilon, p) \rightarrow (0, 0.5)$ (bright yellow region in Fig. 5.1.1 B). This is the region where the observer makes an instantaneous decision after receiving a green light. The green light impacts the expected

5.1. FEEDBACK ENHANCES THE OBSERVER'S INITIAL BELIEF

net reward more, thus the optimal thresholds increase. On average signaling leads to optimal strategies with increased accuracy, whereas probabilistic reward leads to optimal strategies with decreased accuracy.

The increase in initial beliefs leads to more frequent immediate decisions. The ideal observer commits an immediate decision when the initial belief is above threshold, i.e., $\theta \leq |y_{0\pm}^2|$. We denote the optimal thresholds for probabilistic signaling and probabilistic reward by $\theta_s^{\max}$ and $\theta_r^{\max}$, respectively. Thus, the solid (dashed) curve in Fig. 5.1.1 indicates $|y_{0\pm}^2| = \theta_s^{\max}$ ($|y_{0\pm}^2| = \theta_r^{\max}$). In Fig. 5.1.1, the immediate decision occurs for parameter values in the region above the curve for the probabilistic signaling (solid) and on the left region for the probabilistic reward (dashed). In the probabilistic reward case, the observer always makes an immediate decision at $p = 0.5$ because the optimal threshold $\theta_r^{\max}(\epsilon, 0.5) = 0$ for any $0 \leq \epsilon \leq 0.5$. When the reward is given for correct choice with probability $p = 0.5$, the probability of correct response does not contribute to ENR, so setting threshold to be 0 improves the decision time and ENR overall.

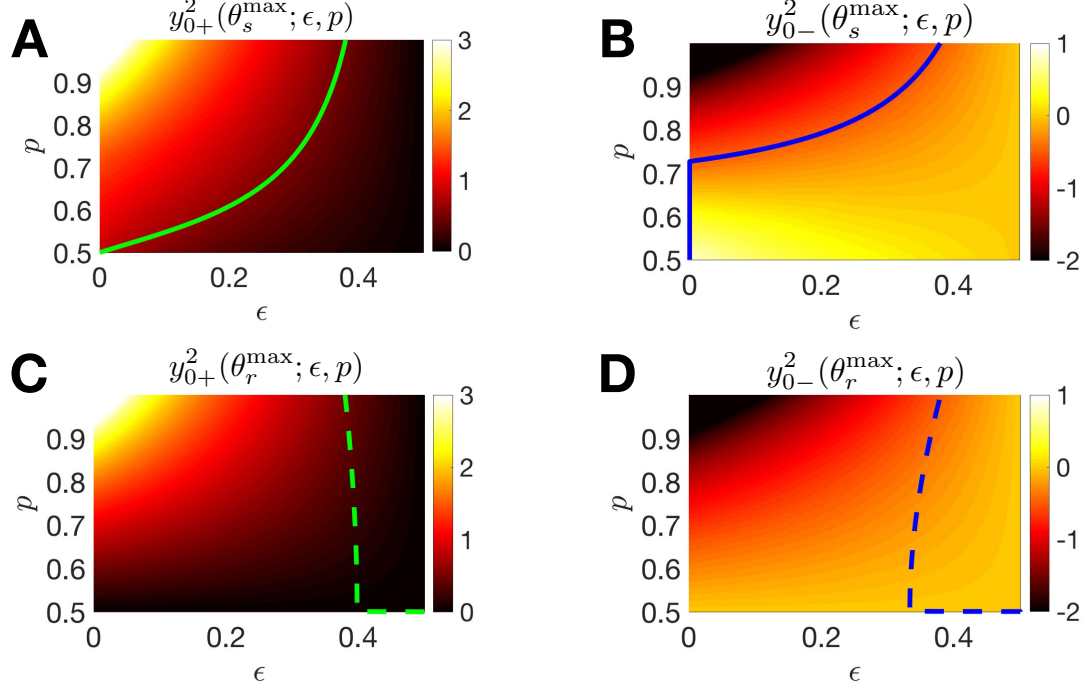


Fig. 5.1.1: Color plots of initial belief on trial 2, $y_{0\pm}^2$, given by light signals: **A.** green light, **B.** red light; and by reward feedback: **C.** reward, **D.** no reward. The optimal policy is for constant thresholds across trials. The curves indicate the parameter values (ϵ, p) at which $\theta_s^{\max} = |y_{0\pm}^2|$, or $\theta_r^{\max} = |y_{0\pm}^2|$, i.e., decisions occur immediately. The solid (dashed) curve corresponds to the case of probabilistic signaling (probabilistic reward) while the green (blue) color indicates positive (negative) feedback.

Positive feedback yields a larger region of immediate decisions because the observer's belief and the feedback agree. When feedback is unreliable, $p \rightarrow 0.5$, the observer can still react instantly when the environment is likely to stay constant, $\epsilon \rightarrow 0$. Negative feedback does not agree with the internal belief, so the initial bias on the next trial is pushed toward the opposite threshold. Unless the probability of correct feedback is high, the information from negative feedback is not sufficient to lead to an immediate decision on trial 2.

5.2. PROBABILISTIC FEEDBACK REDUCES MEAN DECISION TIME AND INCREASES PROBABILITY OF CORRECT RESPONSES ON TRIAL 2

The immediate-decision regions are smaller in the case of probabilistic signaling. This is because there is more incentive to ensure a correct response in the case of probabilistic signaling as opposed to probabilistic reward. In the case of probabilistic signaling, the observer can still be rewarded after they receive negative feedback. If the feedback is not reliable, the observer will need to accumulate information to increase their accuracy. In contrast, when the probability of receiving a reward for being correct is low, the observer will react faster since a correct response does not ensure a reward. In this case, there is more to be gained by simply reducing the time to a decision to increase overall performance on the task.

5.2 Probabilistic Feedback Reduces Mean Decision Time and Increases Probability of Correct Responses on Trial 2

Because the observer knows the probability of receiving correct feedback, they can estimate the probability of receiving positive ($s_1 = +1$) or negative ($s_1 = -1$) feedback on each trial by conditioning on both outcomes of the same trial

$$\begin{aligned} P(s_1 = +1) &= P(s_1 = +1|H^1 = H_{\pm}, d_1 = \pm 1)P(d_1 = \pm 1|H^1 = H_{\pm}) \\ &\quad + P(s_1 = +1|H^1 = H_{\mp}, d_1 = \pm 1)P(d_1 = \pm 1|H^1 = H_{\mp}) \\ &= (2p - 1)c_1 + (1 - p). \end{aligned} \tag{5.2.1}$$

5.2. PROBABILISTIC FEEDBACK REDUCES MEAN DECISION TIME AND INCREASES PROBABILITY OF CORRECT RESPONSES ON TRIAL 2

Similarly,

$$P(s_1 = -1) = (1 - 2p)c_1 + p. \quad (5.2.2)$$

Note in the limit $p \rightarrow 0.5$, $P(s_1 = \pm 1) \rightarrow 0.5$, so the signals are decoupled from the probability of a correct response. On the other hand, as $p \rightarrow 1$, the feedback perfectly corresponds to whether the observer was right or wrong, i.e., $P(s_1 = +1) = c_1$ and $P(s_1 = -1) = 1 - c_1$.

5.2.1 Probability of Being Correct Depends on Reliability of Feedback

Feedback is received only after a decision is made. We assume that both options are equally likely at the beginning of trial 1, so the observer has neither bias nor information about the correct choice, and trial 1 is equivalent to the case of no feedback. Therefore,

$$c_1 = \frac{1}{1 + e^{-\theta/D}}.$$

Assuming the feedback is positive $s_1 = +1$ and the observer makes decision

5.2. PROBABILISTIC FEEDBACK REDUCES MEAN DECISION TIME AND INCREASES PROBABILITY OF CORRECT RESPONSES ON TRIAL 2

$d_1 = +1$, LLR on trial 2 if the initial belief is below threshold is

$$\begin{aligned} \frac{P(H^2 = H_+ | d_2 = +1, s_1 = +1)}{P(H^2 = H_- | d_2 = +1, s_1 = +1)} &= e^{\theta/D} \\ \frac{c_2 | (s_1 = +1)}{1 - c_2 | (s_1 = +1)} &= e^{\theta/D} \\ c_2 | (s_1 = +1) &= \frac{1}{1 + e^{-\theta/D}}. \end{aligned} \quad (5.2.3)$$

If the initial belief is above threshold, i.e., $y_{0+}^2 > 0$, then the observer makes decision $d_2 = +1$ and

$$\begin{aligned} \frac{P(H^2 = H_+ | d_2 = +1, s_1 = +1)}{P(H^2 = H_- | d_2 = +1, s_1 = +1)} &= e^{y_{0+}^2/D} \\ \frac{c_2 | (s_1 = +1)}{1 - c_2 | (s_1 = +1)} &= e^{y_{0+}^2/D} \\ c_2 | (s_1 = +1) &= \frac{1}{1 + e^{-y_{0+}^2/D}}. \end{aligned} \quad (5.2.4)$$

Otherwise, when $y_{0+}^2 < 0$, then the decision is $d_2 = -1$ and

$$\begin{aligned} \frac{P(H^2 = H_+ | d_2 = -1, s_1 = +1)}{P(H^2 = H_- | d_2 = -1, s_1 = +1)} &= e^{y_{0+}^2/D} \\ \frac{1 - c_2 | (s_1 = +1)}{c_2 | (s_1 = +1)} &= e^{y_{0+}^2/D} \\ c_2 | (s_1 = +1) &= \frac{1}{1 + e^{y_{0+}^2/D}}. \end{aligned} \quad (5.2.5)$$

In general, the correct probability given the previous signal is:

$$c_2 | (s_1 = \pm 1) = \frac{1}{1 + e^{-\max(\theta, |y_{0\pm}^2|)/D}}. \quad (5.2.6)$$

5.2. PROBABILISTIC FEEDBACK REDUCES MEAN DECISION TIME AND INCREASES PROBABILITY OF CORRECT RESPONSES ON TRIAL 2

Another way to write Eq. (5.2.6) is

$$c_2|(s_1 = +1) = \begin{cases} \frac{1}{1+e^{-\theta/D}} & : \theta > |y_{0+}^2|, \\ \frac{c_1(p-\epsilon)+(1-p)\epsilon}{c_1(2p-1)+(1-p)} & : \theta \leq |y_{0+}^2|, \end{cases} \quad (5.2.7)$$

and

$$c_2|(s_1 = -1) = \begin{cases} \frac{1}{1+e^{-\theta/D}} & : \theta > |y_{0-}^2|, \\ \frac{c_1(\epsilon-p)+p(1-\epsilon)}{c_1(1-2p)+p} & : \theta \leq |y_{0-}^2|. \end{cases} \quad (5.2.8)$$

The probability of correct response on trial 2 is marginalized on both positive and negative signals

$$c_2 = \begin{cases} c_1 & : \theta > |y_{0+}^2| > |y_{0-}^2|, \\ (1-2p)c_1^2 + (2p-\epsilon)c_1 + (1-p)\epsilon & : |y_{0+}^2| \geq \theta > |y_{0-}^2|, \\ p(1-2\epsilon) + \epsilon & : |y_{0+}^2| \geq |y_{0-}^2| \geq \theta, \end{cases} \quad (5.2.9)$$

where $0.5 \leq p - \epsilon \leq p(1 - 2\epsilon) + \epsilon \leq p \leq 1$.

5.2. PROBABILISTIC FEEDBACK REDUCES MEAN DECISION TIME AND INCREASES PROBABILITY OF CORRECT RESPONSES ON TRIAL 2

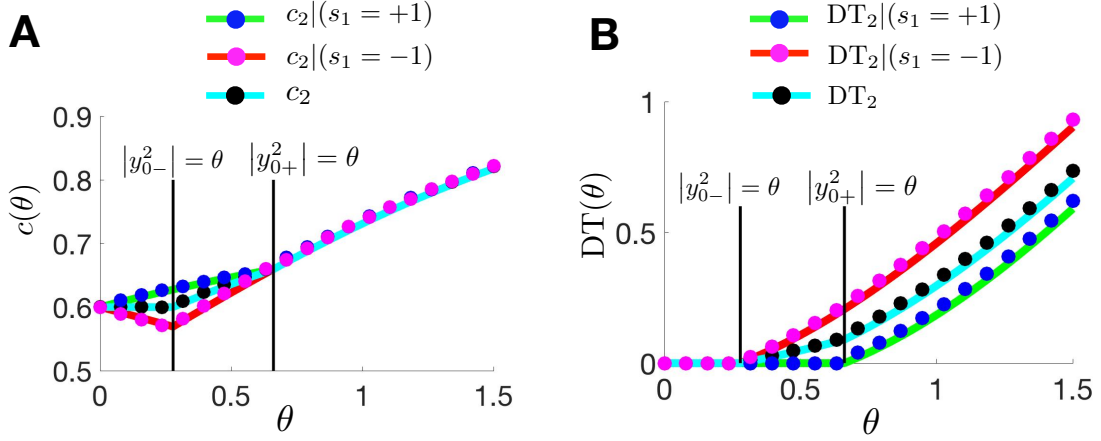


Fig. 5.2.1: Here, the probability of environmental change is $\epsilon = 0.25$ and probability of correct feedback is $p = 0.7$. The stochastic simulations (dots) are compared to theoretical predictions (solid lines). Feedback improves the probability of being correct in the case of quicker decisions. Positive feedback (green curves) leads to both more accurate and speedier decisions as compared to decisions following negative feedback (red curves). This is because most positive feedback follows a correct decision. **A.** Probability of correct responses. **B.** Mean decision time on trial 2.

When both initial beliefs for positive and negative feedback are below threshold, i.e., $\theta > |y_{0+}^2| > |y_{0-}^2|$, feedback does not enhance the correct probability as $c_2 = c_2|(s_1 = \pm 1) = 1/(1 + e^{-\theta/D})$. (the region on the right of $|y_{0-}^2| = \theta$ in Fig. 5.2.1 A). As threshold decreases, the probability being correct c_2 only depends on change rate ϵ and probability of correct feedback p , as initial beliefs for both types of feedback are above threshold. For instance, if the environment does not change, $\epsilon = 0$, and the feedback is perfect, $p = 1$, then the observer is always correct on trial 2 with probability $c_2 = 1$. On the other hand, if the feedback is perfect, $p = 1$ and there is no correlation, $\epsilon = 0.5$, then the accuracy is $c_2 = 0.5$, which is as good as guessing. For high thresholds, feedback does not improve accuracy (region on the right of $|y_{0+}^2| = \theta$ in Fig 5.2.1 A). Because the observer has

5.2. PROBABILISTIC FEEDBACK REDUCES MEAN DECISION TIME AND INCREASES PROBABILITY OF CORRECT RESPONSES ON TRIAL 2

more time to accumulate information, they rely more on their internal belief. As thresholds decrease, the positive feedback has a greater impact on accuracy, and thus it improves the total probability of being correct. Interestingly, correct probability given by negative feedback decreases as thresholds increase for $|y_{0-}^2| \geq \theta$ (see Fig. 5.2.1 A). Recall that the negative feedback pushes the initial belief to the opposite threshold. Since the environment is not static and the feedback is not perfect, the initial belief is more likely to exit the wrong decision boundary. Without any subsequent evidence accumulation, the accuracy will decrease. Therefore, setting higher thresholds does not improve accuracy.

5.2.2 Feedback Decreases Mean Decision Time in an Optimal Observer

Now we can compute the average time until a decision on trial 2, assuming the environment on trial 2 is $H^2 = H_+$ and feedback on trial 1 is positive, $s_1 = +1$:

$$DT_2|(H^2 = H_+, s_1 = +1) = \sum_{k=\pm 1} T(k \cdot y_{0+}^2; \theta). \quad (5.2.10)$$

Thus, the mean decision time on trial 2 given by positive/negative feedback is

$$\begin{aligned} DT_2|(H^2 = H_+, s_1 = +1) = \frac{1}{P(s_1 = +1)} & \left[T(y_{0+}^2; \theta_2) [p(1 - \epsilon)c_1 + (1 - q)\epsilon(1 - c_1)] \right. \\ & \left. + T(-y_{0+}^2; \theta_2) [(1 - q)(1 - \epsilon)(1 - c_1) + p\epsilon c_1] \right], \end{aligned} \quad (5.2.11)$$

5.2. PROBABILISTIC FEEDBACK REDUCES MEAN DECISION TIME AND INCREASES PROBABILITY OF CORRECT RESPONSES ON TRIAL 2

and

$$\begin{aligned} \text{DT}_2|(H^2 = H_+, s_1 = -1) = \frac{1}{P(s_1 = -1)} & \left[T(y_{0-}^2; \theta_2) [(1-p)(1-\epsilon)c_1 + q\epsilon(1-c_1)] \right. \\ & \left. + T(-y_{0-}^2; \theta_2) [q(1-\epsilon)(1-c_1) + (1-p)\epsilon c_1] \right]. \end{aligned} \quad (5.2.12)$$

Substituting the mean first passage time from Eq. (2.2.3) into the unconditional mean decision time equations, we obtain

$$\text{DT}_2|(s_1 = +1) = \begin{cases} \theta \cdot \frac{1-e^{-\theta/D}}{1+e^{-\theta/D}} - y_{0+}^2 (1-2\epsilon) \frac{c_1-(1-p)}{(2p-1)c_1+(1-p)} & : \theta > |y_{0+}^2|, \\ 0 & : \theta \leq |y_{0+}^2|, \end{cases} \quad (5.2.13)$$

and

$$\text{DT}_2|(s_1 = -1) = \begin{cases} \theta \cdot \frac{1-e^{-\theta/D}}{1+e^{-\theta/D}} - y_{0-}^2 (1-2\epsilon) \frac{c_1-p}{(1-2p)c_1+p} & : \theta_2 > |y_{0-}^2|, \\ 0 & : \theta_2 \leq |y_{0-}^2|. \end{cases} \quad (5.2.14)$$

Concerning the response times on trial 2, there are 3 possibilities: (1) both initial beliefs are below threshold and the observer takes a nonzero time to deliberate in either case, (2) the initial belief for positive feedback is above threshold and the observer responds immediately while initial belief for negative feedback is below threshold and leads to further evidence accumulation, and (3) both initial beliefs are above thresholds (look at three regions separated by dashed lines in Fig. 5.2.1).

5.2. PROBABILISTIC FEEDBACK REDUCES MEAN DECISION TIME AND INCREASES PROBABILITY OF CORRECT RESPONSES ON TRIAL 2

The unconditional mean decision time is therefore a piecewise function

$$\begin{aligned}
 DT_2|(H^2 = H_+) &= \sum_{k=\pm 1} DT_2|(H^2 = H_+, s_1 = k)P(s_1 = k) \\
 &= \begin{cases} DT_1 - (1 - 2\epsilon)[y_{0+}^2(p - 1 + c_1) + y_{0-}^2(c_1 - p)] & : \theta > |y_{0+}^2| > |y_{0-}^2|, \\
 ((1 - 2p)c_1 + p)DT_1 - y_{0-}^2(1 - 2\epsilon)(c_1 - p) & : |y_{0+}^2| \geq \theta > |y_{0-}^2|, \\
 0 & : |y_{0+}^2| \geq |y_{0-}^2| \geq \theta. \end{cases}
 \end{aligned} \tag{5.2.15}$$

For high thresholds, there is an increase in mean decision time given by negative feedback. Negative feedback propels the observer's belief away from the previously crossed threshold, so unless it pushes it very close to the other threshold it will take longer for LLR to reach the decision boundary than in the positive feedback case. The mean decision time in the region on the left of $|y_{0-}^2|$ is 0, meaning that the negative feedback is informative enough to trigger a decision without accumulating evidence (see Fig. 5.2.1). In summary, both positive feedback and negative feedback help improve accuracy and lead to faster decisions. This result is consistent with the trends in Fig. 1.2.1. A more interesting question is how the optimal observer changes their strategy for different types of probabilistic feedback. We address this question next by computing reward functions.

5.3 Optimizing Expected Net Reward for Probabilistic Signaling and Probabilistic Reward

A major difference between the contexts of probabilistic signaling and probabilistic reward is the reward schedule. In the first case, a cumulative reward for each correct choice is given at the end of the trial sequence while in the second case, a reward is given probabilistically at the end of each trial. This difference affects the computation of expected reward, primarily because correct choices are not all rewarded and incorrect choices are sometimes rewarded in the case of probabilistic reward. In this section, we identify how these differences shape the strategy and response features using the ENR function as our performance measure.

5.3.1 Probabilistic Signaling

In this case, reward is deterministic, so we define ENR as follows:

$$\text{ENR}_{1:2} = c_1 + \gamma c_2 - \text{cost} \cdot [\text{DT}_1 + T_D + \gamma(\text{DT}_2 + T_D)], \quad (5.3.1)$$

where c_1 , c_2 , DT_1 , and DT_2 are defined from the last sections, and γ is discount rate.

In the color plot of optimal thresholds for different values of ϵ and p , there are 3 regions (see Fig. 5.3.1). In the region below the green curve, both initial beliefs are below thresholds. Recall that feedback does not matter in this region as the ideal observer relies more on their internal belief. The optimal thresholds

5.3. OPTIMIZING EXPECTED NET REWARD FOR PROBABILISTIC SIGNALING AND PROBABILISTIC REWARD

increase as the change probability ϵ gets closer to 0. This result is consistent with the results in chapter 4 (see Fig. 4.2.1 A-C). In the region above the blue curve, both initial beliefs are above thresholds. In this region, the ideal observer relies solely on feedback. This is also an immediate-decision region because there is no need to accumulate information on trial 2. The optimal thresholds are the lowest at point $(\epsilon, p) = (0, 1)$ because the feedback is perfect and is perfectly transferred forwards to trial 2. In contrast to the region below the green curve, as ϵ increases, the optimal thresholds increase as well.

The region between the blue and green curves is our interest as the observer makes an immediate decision on trial 2 after a green signal but must accumulate evidence after a red signal. We call such region the “conflict” region as this shows the conflicting evidence accumulation strategies provoked by receiving positive versus negative feedback and is reminiscent of results from prior experimental work [92].

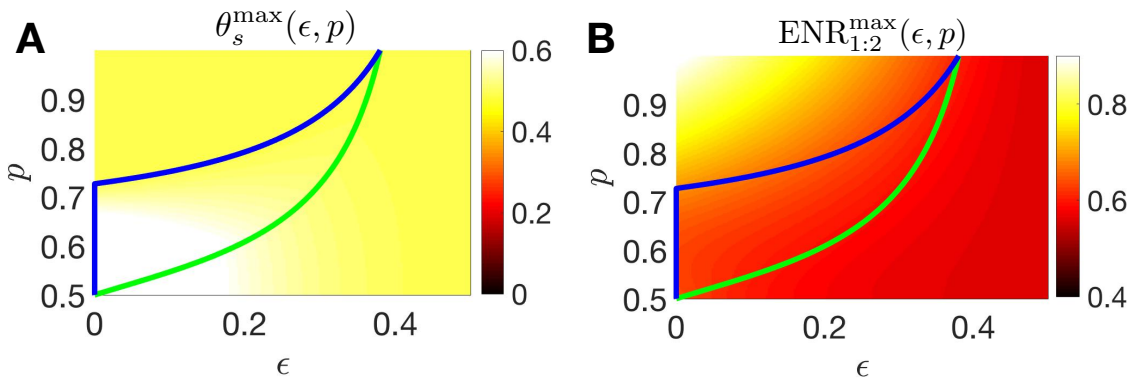


Fig. 5.3.1: In the case of probabilistic signaling, color plots of **A.** Optimal threshold θ and **B.** the maximum ENR. The green (blue) curve denotes where $|y_{0+}^2| = \theta_s^{\max}$ ($|y_{0-}^2| = \theta_s^{\max}$), so in the region (ϵ, p) to the left of these curves, responses are immediate on trial 2.

5.3. OPTIMIZING EXPECTED NET REWARD FOR PROBABILISTIC SIGNALING AND PROBABILISTIC REWARD

Positive feedback always leads to a stronger subsequent initial bias than in the case of negative feedback. As such decision times on trial 2 are always longer (or equal) following a red signal than following a green signal, and the region of immediate decisions is larger in the case of the green signal (Fig. 5.3.2). In the immediate-decision region, the conditional mean decision time is 0, and the conditional correct probability given by green signal increases as probability of correct signal p increases. However, in the evidence accumulation region, there is a small area where the decision time on trial 2 following a red signal varies non-monotonically with the feedback parameter p . This is due to the fact that a weak signal will tend to drive the initial belief on trial 2 away from the closest threshold (when ϵ is small), but as p is increased further, the red signal drives the initial belief on trial 2 closer to the opposite threshold, triggering quicker decisions for the opposite choice from before. Increasing p and decreasing ϵ increases the total correct probability and decreases the mean decision time on trial 2 as expected.

5.3. OPTIMIZING EXPECTED NET REWARD FOR PROBABILISTIC SIGNALING AND PROBABILISTIC REWARD

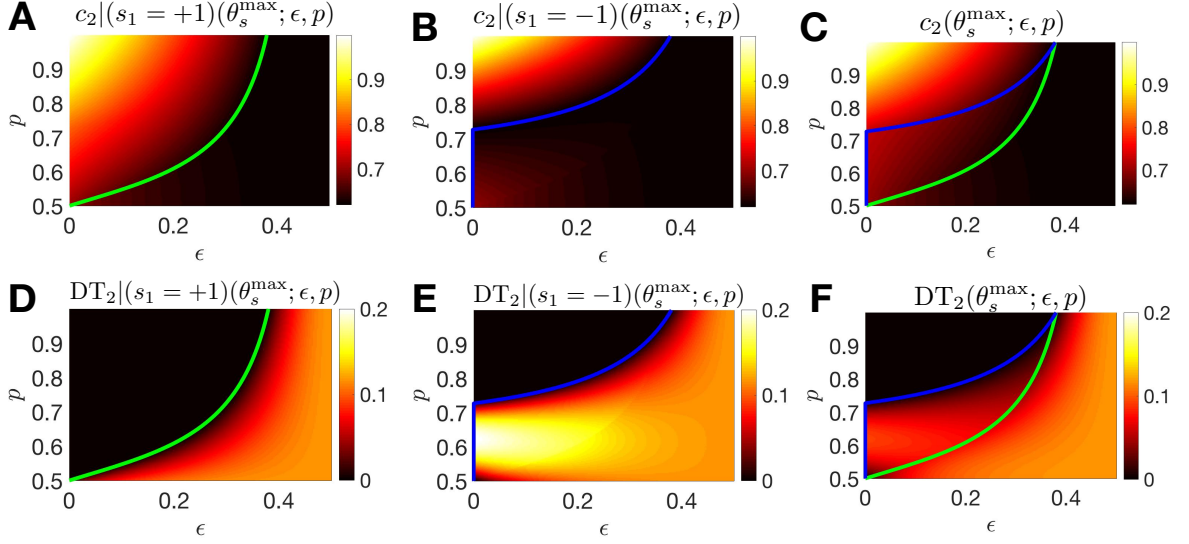


Fig. 5.3.2: Color plots of probability of correct response and mean decision time on trial 2 in the context of probabilistic signaling with parameter p . The green (blue) solid curve indicates where $|y_{0+}^2| = \theta_s^{\max}$ ($|y_{0-}^2| = \theta_s^{\max}$). Optimal decision thresholds $\theta_s^{\max}$ are found by optimizing Eq. (5.3.1). The regions above the curves are “immediate-decision” regions. Correct probability on trial 2 conditioned on **A.** green light, $c_2|(s_1 = +1)$, **B.** red light, $c_2|(s_1 = -1)$, and **C.** probability being correct c_2 . **D.** Decision time on trial 2 conditioned on **D.** green light, $DT_2|(s_1 = +1)$; **E.** conditioned on red light, $DT_2|(s_1 = -1)$, and **F.** mean decision time on trial 2, DT_2 .

5.3.2 Probabilistic Reward

When rewards are given probabilistically, the observer receives rewards for correct choices with probability p and rewards for incorrect choices with probability $1 - p$. Therefore, the ENR is defined as

$$\begin{aligned}
 \text{ENR}_{1:2} &= pc_1 + (1 - p)(1 - c_1) + \gamma[pc_2 + (1 - p)(1 - c_2)] \\
 &\quad - \text{cost} \cdot [\text{DT}_1 + T_D + \gamma(\text{DT}_2 + T_D)] \\
 &= (2p - 1)(c_1 + \gamma c_2) + (1 - p)(1 + \gamma) - \text{cost} \cdot [\text{DT}_1 + T_D + \gamma(\text{DT}_2 + T_D)].
 \end{aligned} \tag{5.3.2}$$

where c_1 and DT_1 are defined in Eq. (2.3.5) and (2.2.4), c_2 and DT_2 are defined in Eq. (5.2.9) and (5.2.15).

In Fig. 5.3.3 A, the probabilistic reward yields different optimal thresholds comparing to the probabilistic signaling. The regions on the left of the green (blue) dashed curves are immediate-decision regions, where feedback plays an important role in decision-making. In contrast to probabilistic signaling case, the negative feedback matters more as the immediate-decision region for negative feedback is larger. For small values of p , the probability of receiving a reward for correct choice is low, so the observer prioritizes speed over accuracy since their probability of being correct matter little to their ENR, so they set a low thresholds to make quicker decisions. For instance, when $p = 0.5$, the observer only receives rewards half of the time they are correct and half the time when they are incorrect. As such, it is best to guess immediately, so the optimal threshold is set to be 0, regardless of the value of ϵ . As p increases, correct choices are more valuable,

5.3. OPTIMIZING EXPECTED NET REWARD FOR PROBABILISTIC SIGNALING AND PROBABILISTIC REWARD

encouraging the observer to spend more time accumulating evidence.

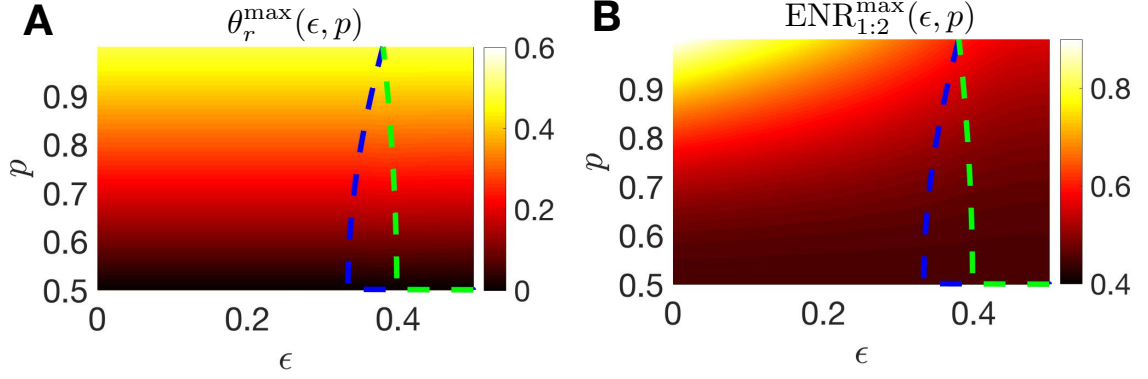


Fig. 5.3.3: In the case of probabilistic reward, color plots of **A.** optimal threshold $\theta_r^{\max}$ and **B.** maximum $\text{ENR}_{1:2}^{\max}$. The green (blue) dashed curve is $|y_{0+}^2| = \theta_r^{\max}$ ($|y_{0-}^2| = \theta_r^{\max}$). Note that the regions in which an instantaneous decision is triggered on trial 2 (left of curves) is larger than in the case of probabilistic signaling.

Similar to the signal-based feedback, the accuracy increases when the feedback is more reliable ($p \rightarrow 1$) (see Fig. 5.3.4 A, B, and C) because strong and reliable feedback enhances the observer's bias and leads to improvement in accuracy. When the environment is less correlated ($\epsilon \rightarrow 0.5$), the observer does not have a strong bias at the beginning of trial 2, so they need to spend time to accumulate evidence. This shows in Fig. 5.3.4 D, E, and F, where the mean decision time increases as $\epsilon \rightarrow 0.5$.

For any pairs of (ϵ, p) on the immediate-decision region, from Eq. (5.2.9), the probability of correct response on trial 2 depends on ϵ and p , particularly, $c_2 = p(1 - 2\epsilon) + \epsilon$. When $p \rightarrow 1$, i.e., perfect feedback, and $\epsilon \rightarrow 0$, the correct probability $c_2 \rightarrow 1$, which is the highest accuracy the observer can attain. Thus, the observer can react immediately while their accuracy increases when the feedback is reliable and the environment is likely to stay constant, i.e., $(\epsilon, p) \rightarrow (0, 1)$.

5.3. OPTIMIZING EXPECTED NET REWARD FOR PROBABILISTIC SIGNALING AND PROBABILISTIC REWARD

Reducing the reliability of feedback reduces the accuracy.

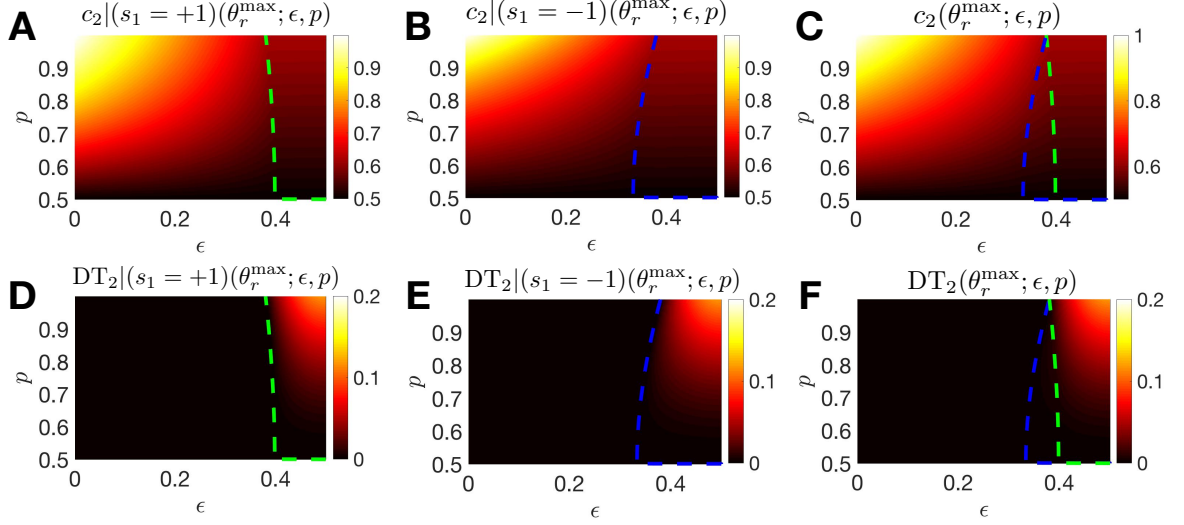


Fig. 5.3.4: **A, B, C.** Color plots of probability of correct response and **D, E, F.** mean decision time on trial 2 following a probabilistic reward on trial 1. The green (blue) curve indicates the collision between initial belief for positive (negative) feedback and threshold $y_{0+}^2 = \theta_r^{\max}$ ($y_{0-}^2 = \theta_r^{\max}$). The positive (negative) feedback is defined as the observer receive reward (no reward). Optimal thresholds $\theta_r^{\max}$ are found from optimizing Eq. (5.3.2). **A, D.** The conditional correct probability and decision time given by positive feedback. **B, E.** The conditional correct probability and decision time given by negative feedback. **C, F.** Total correct response and mean decision time.

Reward-based feedback yields larger immediate-decision regions (regions on the left of the curves) compared to signal-based feedback. The negative feedback shares a large immediate-decision region with positive feedback for probabilistic reward. There is not much conflict between positive reward-based feedback and negative reward-based feedback. For small values of p , the probability of receiving reward (no reward) for correct choice (incorrect choice) is low, so regardless of

types of feedback, the observer will react quickly. An observer receiving reward-based feedback focuses on obtaining a higher expected reward by making quicker decisions, whereas observers receiving signal-based feedback can focus more on being correct since rewards are always given for correct choices. Overall, signal-based feedback yields higher expected net reward.

Remark: In the results displayed for our ENR maximization analysis (Fig. 5.3.1, 5.3.2, 5.3.3 and 5.3.4), we have fixed $\gamma = 0.8$, $\text{cost} = T_D = 0.5$, and $D = 1$.

5.4 Summary

In this chapter, we introduce probabilistic signaling and probabilistic reward as feedback. Optimal observers receiving probabilistic signaling opt for higher decision thresholds than those receiving probabilistic reward. If probability of receiving reward for correct decisions is low, there is no need to keep past information as improving accuracy only modestly increases the expected net reward overall. We find that reward-based feedback yields a larger immediate-decision region and a smaller region of conflict between positive and negative feedback. For a low probability of correct signal-based feedback, the observer is certain that they receive reward for correct choice, so evidence-accumulation process is needed. On the other hand, for a low probability of correct reward-based feedback, the observer does not need to waste time accumulating evidence as the probability of receiving reward for being correct is low. This explains why the conflict region is larger in the case of signal-based feedback. We are also interested in how well the predictions of our model match human subject reaction times and accuracy in

5.4. SUMMARY

2AFC correlated tasks. In the next chapter, we compare our no-feedback model to experimental data trends.

Discussion

6.1 The Optimal Observer Assumes Incorrect Change Probability

Our modeling work demonstrates that there can be a tension between the information provided from evidence collected on a previous trial and feedback in 2AFC tasks with correlated trials. Correlations in the environment make it ideal to value evidence received in earlier trials more than in later trials, and thus deliberations should be longer in the first few trials. When considering the effects of feedback, we find that the evidence discounting effects of environmental state changes can be counteracted by feedback received at the ends of trials. As such, the evidence that an agent has in favor of an option at the beginning of a trial can be more than they had at the end of the last trial prior to feedback. We also show here that the typical idiosyncratic repetition biases commonly observed in human

6.1. THE OPTIMAL OBSERVER ASSUMES INCORRECT CHANGE PROBABILITY

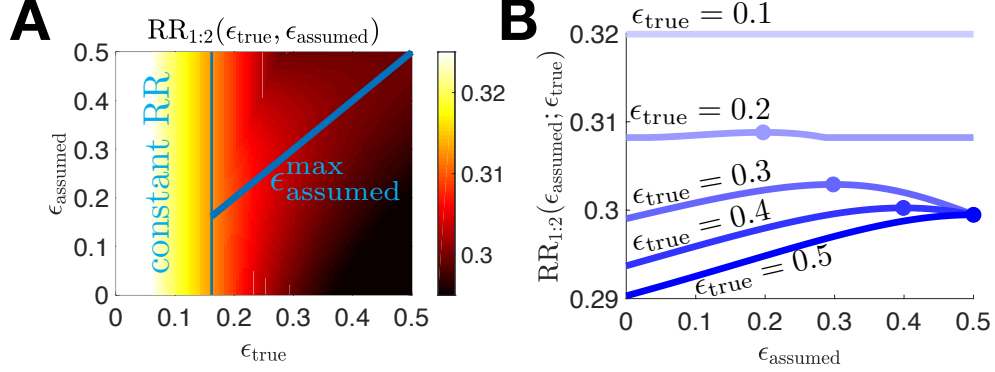


Fig. 6.1.1: The optimal RR for a two trial task as a function of the true ϵ_{true} and the assumed $\epsilon_{assumed}$ switching rates. Eq. (3.2.1) is maximized by tuning (θ_1, θ_2) while fixing $(\epsilon_{true}, \epsilon_{assumed})$. **A.** For slow environments (small ϵ_{true}), the RR obtains the same maximum regardless of the choice of $\epsilon_{assumed} < 0.5$, while for larger ϵ_{true} the optimal choice $\epsilon_{assumed}^{max} = \epsilon_{true}$. **B.** The RR varies weakly with $\epsilon_{assumed}$ about the maximum RR (dots) for each ϵ_{true} .

subjects emerge naturally in a normative model that assumes environmental correlations.

We now 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 a series of 2AFC trials [32], even when the trials are uncorrelated [17, 35]. For instance, when the state is the same on two adjacent trials, the probability of a correct response on the second trial increases [34], and the response time decreases [45].

In our model, if the assumed change rate of the environment does not match the true change rate, $\epsilon_{assumed} \neq \epsilon_{true}$, performance can decrease, compared to $\epsilon_{assumed} = \epsilon_{true}$. However, this decrease is not particularly severe. We studied the maximal RR from Eq. (3.2.1) for each $(\epsilon_{true}, \epsilon_{assumed})$ while tuning (θ_1, θ_2) (Fig. 6.1.1). For ϵ_{true} small, the RR is constant across all $\epsilon_{assumed}$ because deciding

6.1. THE OPTIMAL OBSERVER ASSUMES INCORRECT CHANGE PROBABILITY

immediately in trial 2 is always optimal in this case (Fig. 6.1.1A). For larger ϵ_{true} , taking $\epsilon_{\text{assumed}} = \epsilon_{\text{true}}$ is optimal as expected, but the RR depends only weakly on $\epsilon_{\text{assumed}}$ (Fig. 6.1.1B). This suggests that subjects pay little penalty for implementing a model with a mistuned $\epsilon_{\text{assumed}}$. Thus experimentally observed response trends may arise partly due to misestimation of the transition rate ϵ_{true} and the assumption that trials are dependent, even if they are not [17, 97].

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

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 coherence* (See Fig. 6.3.1A). 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. (2.3.4), and the initial decision variable, y_0^j , is given by Eq. (3.1.4) using the threshold θ and the assumed transition rate, $\epsilon_{\text{assumed}}$. 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 $\epsilon_{\text{assumed}} = \epsilon^{\text{true}}$.

*We define coherence as the drift, $g^j \in \pm 1$, in trial j divided by the noise diffusion coefficient D [8, 38].

Fig. 6.3.1A 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. (4.2.3). 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 Repetition-Dependence of Correct Probabilities and Decision Times

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), $c_j|R := c_j|(H^j = H^{j-1} = H_{\pm})$

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

6.2. REPETITION-DEPENDENCE OF CORRECT PROBABILITIES AND DECISION TIMES

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 $\theta > 0$

6.2. REPETITION-DEPENDENCE OF CORRECT PROBABILITIES AND DECISION TIMES

and $\epsilon \in [0, 0.5)$. To begin, note that

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

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 $F'(y) = [\frac{y}{D} \cosh \frac{y}{D} - \sinh \frac{y}{D}] / y^2 \geq 0$ 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 $c_j|A$ as well as $\pi_\theta(y)$ described by Eq. (2.3.4) 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

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

6.3 Our Model Is Consistent With Experimental Trend

The adaptive DDM also exhibits repetition bias in decision times [40]. 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 time is increased if the state is different (alternation, see section 6.2). This occurs for optimal ($\epsilon_{\text{assumed}} = \epsilon_{\text{true}}$) and suboptimal ($\epsilon_{\text{assumed}} \neq \epsilon_{\text{true}}$) models as long as $\epsilon_{\text{assumed}} \in [0, 0.5)$. Furthermore, the rate of correct responses increases for repetitions and decreases for alternations (see section 6.2). 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 [17, 40, 45], we propose that such repetition biases in the case of uncorrelated trials could be the result of the subject assuming $\epsilon_{\text{assumed}} \neq 0.5$, even when $\epsilon_{\text{true}} = 0.5$.

Following [40], we denote that three adjacent trial states are composed of repetitions (R), and alternations (A) [17, 40]. Repetition-repetition (RR)[†] corresponds

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

6.3. OUR MODEL IS CONSISTENT WITH EXPERIMENTAL TREND

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 section 6.2. Here we assume that $\epsilon_{\text{assumed}} = \epsilon_{\text{true}}$ for simplicity, although it is likely that subjects often use $\epsilon_{\text{assumed}} \neq \epsilon_{\text{true}}$ as shown in [17, 40].

6.3. OUR MODEL IS CONSISTENT WITH EXPERIMENTAL TREND

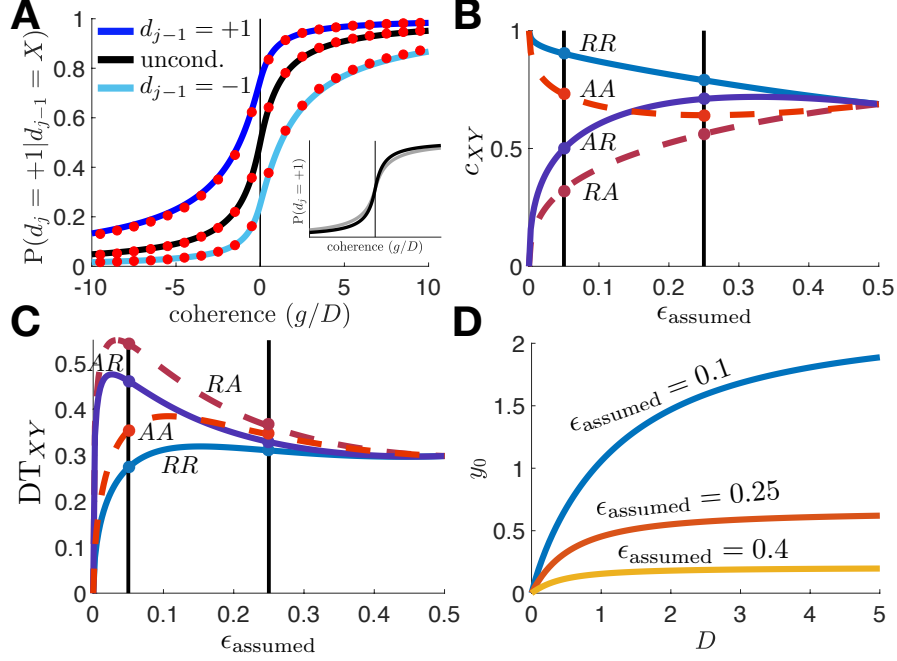


Fig. 6.3.1: Repetition bias trends in the adaptive DDM. **A.** The psychometric functions in trial j conditioned on d_{j-1} shows that when $d_{j-1} = +1$, the fraction of $d_j = +1$ responses increases (top curve). When $d_{j-1} = -1$, the fraction of $d_j = +1$ responses decreases (bottom curve). The weighted average of these two curves gives the unconditioned psychometric function (middle curve). Coherence g/D is defined for drift $g \in \pm 1$, noise diffusion coefficient D , and assumed change rate $\epsilon_{\text{assumed}} = 0.25$. 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 $\epsilon_{\text{assumed}}$, 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 $\epsilon_{\text{assumed}}$ is increased, since the effects of the state two trials back are weakened. Compare dots at $\epsilon_{\text{assumed}} = 0.05$ and $\epsilon_{\text{assumed}} = 0.25$. **C.** Decision time DT_{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 $\epsilon_{\text{assumed}} = 0.1, 0.25, 0.4$. In all cases, we set $T_D = 2$, $\theta = \theta_{\infty}^{\max}$ for all trials, and $D = 1$ if not mentioned.

The probability of a correct response as a function of $\epsilon_{\text{assumed}}$ is shown for all four conditions in Fig. 6.3.1B. Clearly, c_{RR} is always largest, since the ideal

6.3. OUR MODEL IS CONSISTENT WITH EXPERIMENTAL TREND

observer correctly assumes repetitions are more likely than alternations (when $\epsilon_{\text{assumed}} \in [0, 0.5)$).

Note when $\epsilon_{\text{assumed}} = 0.05$, $c_{RR} > c_{AA} > c_{AR} > c_{RA}$ (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 $\epsilon_{\text{assumed}}$, 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 $\epsilon_{\text{assumed}}$, 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 assumed to be quite rare when $\epsilon_{\text{assumed}}$ is small. As $\epsilon_{\text{assumed}}$ is increased, the ordering switches to $c_{RR} > c_{AR} > c_{AA} > c_{RA}$ (e.g., at $\epsilon_{\text{assumed}} = 0.25$), 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, DT_{XY} , conditioned on all possible three-state sequences (see section 6.2). Fig. 6.3.1C shows that, as a function of $\epsilon_{\text{assumed}}$, decision times following two repetitions, DT_{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

6.3. OUR MODEL IS CONSISTENT WITH EXPERIMENTAL TREND

long when the decision variable starts close to the boundary. Also, note that when $\epsilon_{\text{assumed}} = 0.05$, $DT_{RR} < DT_{AA} < DT_{AR} < DT_{RA}$, and the explanation is similar to that for the probabilities of correct responses. When the environment is assumed to change slowly, decisions after two alternations will be quicker because bias in the direction of the correct decision is more likely. The ordering is different at intermediate $\epsilon_{\text{assumed}}$ where $DT_{RR} < DT_{AR} < DT_{AA} < DT_{RA}$.

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 [17, 35, 40]: Sometimes $c_{AA} > c_{AR}$ and $DT_{AA} < DT_{AR}$ whereas sometimes $c_{AA} < c_{AR}$ and $DT_{AA} > DT_{AR}$. However, it appears to be more common that AA sequences lead to slower decision times, $DT_{AA} < DT_{AR}$ (e.g., See Fig. 3 in [40]). This suggests that subjects assume state switches are rare, corresponding to a smaller $\epsilon_{\text{assumed}}$.

Conclusion & Future Work

It has been known for nearly a century that observers take into account previous choices when making decisions [32]. Notably, Yu et al. developed a theory of sequential effects using a Bayesian framework similar to that presented here [97]. This theory was later extended to capture both first (marginal), and second (sequential) effects [45, 49], predicting that decision times should decrease for repeated trials and increase for alternating trials when an observer assumes a correlated environment. However, these studies did not formalize the link between the decision threshold and subsequent initial condition or study the problem of optimizing the reward rate. In addition, modeling studies have used generalized versions of the DDM to account for trends observed in experimental data [34, 40], but have not examined how to optimize a cost function, like the reward rate. Here we have linked the normative model of evidence-accumulation in correlated environments to a decision strategy that optimally trades off speed and accuracy by maximizing an overall reward rate.

The fact that repetitions decrease response times has long been known [28, 76]. Whether this effect is driven by sensory after-effects of the stimuli, or the evidence and resulting responses are still debated [65]. However, there is experimental evidence that past outcomes bias the initial belief on a subsequent trial in human observers [34, 47]. In contrast to our model, experimental studies also show that responses depend on more than just one preceding trial (reviewed in [54], chapter 6.6). Such trends could be due to continuous adaptation of the evidence-discounting process to the statistics of the environment [45, 97]. The experimentally observed increases in response time, and reductions in erroneous responses following a correct response [51], are consistent with a dynamical threshold model. Normative models of the type we suggest also cannot explain negative recency effects, i.e., the observed tendency to alternate choices [44].

To account for information obtained on one trial observers could adjust accumulation speed [22, 72, 90], threshold [8, 22, 40], or their initial belief on the subsequent trial [8, 11]. 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 [11, 34, 63]. On the other hand, the effect of a dynamic threshold across trials diminishes as the number of trials increase: $RR_{1:n}^{\max}$ 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 [5].

We do note, however, that the fact that the optimal variable-threshold model predicts a decrease in decision times may lend some perspective to the finding that subjects tend to increase the speed of decisions as an experiment goes on [25, 67]. Past studies have suggested that subjects successively decrease their decision threshold until they make an error, and then increase the threshold on the next trial. This theory relies on the presence of reward-based feedback, so it would be interesting to examine decision time update strategies for spans of trials that receive no feedback.

It has been shown that subjects performing decision tasks may still accumulate information even after providing a response [68, 77]. Such effects have been modeled with a DDM that continues to evolve following the decision threshold-crossing. We can incorporate these effects into our normative model by defining the initial conditions on trial $n + 1$ using an ending value $y^n(T_n + \Delta T)$ with ΔT determining the time of integration after a threshold, $\pm\theta$, has been crossed: $y^n(T_n + \Delta T) \sim p(y, T_n + \Delta T | y^n(T_n) = \pm\theta)$. Explicit results may still be possible in this case, but the formulas would be considerably more complicated since y_0^{n+1} would be a random variable.

We also assumed that an observer knows the length of the trial sequence, and uses this number to optimize reward rate [61]. 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, $n \sim \text{Geom}(\lambda)$, then

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

This equation has the same form as Eq. (4.2.2), the RR in the case of fixed trial number n , with n replaced by the average $\langle n \rangle = 1/\lambda$. Thus, we expect that as λ decreases, and the average trial number, $1/\lambda$, increases, there is an increase in the RR and larger optimal decision thresholds $\theta^{\max}$ (as in Fig. 4.2.1). 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 4.3.

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_{true} across a sequence of trials, and assume trials are correlated, even when they are not [17, 35, 97]. 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 [69]. This would result in model ideal observers that use the current estimate of the ϵ to set the initial bias on a trial [97]. A similar approach could also allow us to model observers that incorrectly learn,

or make wrong assumptions, about the rate of change.

Experimental evidence strongly suggests that humans assume the world is changing, even when instructed that experimental conditions will remain unchanged [3, 33, 60]. At the same time, humans can use learned statistical structure of the environment to make decisions [47, 49, 58, 97]. Mathematical models that capture human behavior should, therefore, balance adaptivity with stability, to describe how they learn and use the latent structure of the environment. We will pursue such extensions in future work.

Observers may also not have full knowledge of the underlying statistics of the evidence, $f_{\pm}(\xi)$. Several models have been recently developed to describe how a choice and evidence reliability can be inferred simultaneously [21, 46, 56]. Incorporating dynamic updating of the decision threshold to optimize the reward rate can then lead to bounds that collapse or increase depending on the assumed prior distribution over the evidence reliability [24]. One possible extension of our model would be to consider cases in which evidence quality can switch from one trial to the next, potentially causing the noise amplitude and threshold to adapt continuously within a trial in response to new evidence.

Modeling approaches can also point to the neural computations that underlie the decision-making process. Cortical representations of previous trials have been identified [2, 62], 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 [48, 64]. Such synaptic mechanisms may also encode 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 [29, 37]. Ecologically-adapted decision rules may be difficult to train away, especially if they do not adversely impact performance [88].

There is evidence that in humans, outcomes on previous trials bias future decisions. This is the case even when stimuli provide no information about either choice on a trial, so there is technically no correct decision [7]. Indeed the decision made on such trials can be decoded from electroencephalogram recordings prior to stimulus presentation, suggesting it is primarily determined by evidence on the previous trial. It would be interesting to derive the normative evidence-accumulating model in the case in which $H_{\pm}$ trials are interspersed with pure noise trials (with no correct decision). We expect that responses on pure noise trials will exhibit considerable bias in the direction of decisions on the previous trial. Furthermore, we expect that neurophysiologically inspired models that combine aspects of decision-making and working memory would also exhibit such strong biases, as the decision state may be stuck in the attractor state determined by the previous trial, in the absence of new evidence [59].

The normative model we have derived can be used to identify task parameter ranges that will help tease apart the assumptions made by subjects [17, 35, 40].

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.

Numerical Methods

All numerical simulations of the DDM in chapters 3, 4, and 5 were performed using the Euler-Maruyama method with a timestep $dt = 0.0005$, computing each point from 10^5 realizations. To optimize reward rates (RR) and expected net reward (ENR) for probabilistic signaling, we used the Nelder-Mead simplex method (`fminsearch`) in MATLAB. Optimizing ENR for probabilistic reward using `fminsearch` yields negative optimal thresholds. Therefore, we used `fminsearchbnd` by John D’Errico (2020) to restrict the lower bound to be 0. To compute the pairs of (ϵ, p) at the immediate-decision boundary, we used `lsqnonlin` in MATLAB to find the roots of

$$\begin{cases} \frac{\partial \text{ENR}}{\partial \theta} &= 0, \\ \theta_{r,s}^{\max} - |y_{0\pm}^2| &= 0. \end{cases}$$

Bibliography

- [1] Arman Abrahamyan, Laura Luz Silva, Steven C Dakin, Matteo Carandini, and Justin L Gardner. Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences*, 113(25):E3548—E3557, 2016.
- [2] Athena Akrami, Charles D Kopec, Mathew E Diamond, and Carlos D Brody. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature*, 554(7692):368, 2018.
- [3] John R Anderson and Lael J Schooler. Reflections of the environment in memory. *Psychological science*, 2(6):396–408, nov 1991.
- [4] Norman H Anderson. Effect of first-order conditional probability in a two-choice learning situation. *Journal of Experimental Psychology*, 59(2):73–93, 1960.
- [5] Fuat Balci, Patrick Simen, Ritwik Niyogi, Andrew Saxe, Jessica A Hughes, Philip Holmes, and Jonathan D Cohen. Acquisition of decision making criteria: reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics*, 73(2):640–657, 2011.
- [6] Paul Bertelson. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 13(2):90–102, 1961.
- [7] Stefan Bode, David K Sewell, Simon Lilburn, Jason D Forte, Philip L Smith, and Jutta Stahl. Predicting perceptual decision biases from early brain activity. *Journal of Neuroscience*, 32(36):12488–12498, 2012.
- [8] Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D Cohen. The physics of optimal decision making: A formal analysis of models

BIBLIOGRAPHY

- of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4):700–765, 2006.
- [9] Rafal Bogacz, Peter T Hu, Philip J Holmes, and Jonathan D Cohen. Do humans produce the speedaccuracy trade-off that maximizes reward rate? *Quarterly Journal of Experimental Psychology*, 63(5):863–891, may 2010.
- [10] Michael S Brainard and Allison J Doupe. Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews Neuroscience*, 1(1):31–40, 2000.
- [11] Anke Braun, Anne E Urai, and Tobias H Donner. Adaptive history biases result from confidence-weighted Accumulation of Past Choices. *Journal of Neuroscience*, jan 2018.
- [12] Carlos D Brody and Timothy D Hanks. Neural underpinnings of the evidence accumulator. *Current Opinion in Neurobiology*, 37:149–157, 2016.
- [13] Ethan S Bromberg-Martin and Okihide Hikosaka. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, 63(1):119–126, jul 2009.
- [14] Ethan S Bromberg-Martin, Masayuki Matsumoto, and Okihide Hikosaka. Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, 68(5):815–834, 2010.
- [15] Bingni W Brunton, Matthew M Botvinick, and Carlos D Brody. Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128):95–98, 2013.
- [16] Lars Chittka, Peter Skorupski, and Nigel E Raine. Speed–accuracy tradeoffs in animal decision making. *Trends in Ecology & Evolution*, 24(7):400–407, 2009.
- [17] Raymond Y Cho, Leigh E Nystrom, Eric T Brown, Andrew D Jones, Todd S Braver, Philip J Holmes, and Jonathan D Cohen. Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective, & Behavioral Neuroscience*, 2(4):283–299, 2002.
- [18] Aaron J Corcoran and William E Conner. How moths escape bats: predicting outcomes of predatorprey interactions. *The Journal of Experimental Biology*, 219(17):2704–2715, sep 2016.

- [19] Eveline A Crone, K Richard Ridderinkhof, Mijkje Worm, Riek J M Somsen, and Maurits W Van Der Molen. Switching between spatial stimulus-response mappings: A developmental study of cognitive flexibility. *Developmental Science*, 7(4):443–455, 2004.
- [20] Stanislas Dehaene and Mariano Sigman. From a single decision to a multi-step algorithm. *Current Opinion in Neurobiology*, 22(6):937–945, 2012.
- [21] Sophie Deneve. Making decisions with unknown sensory reliability. *Frontiers in Neuroscience*, 6(JUN):1–13, 2012.
- [22] Adele Diederich and Jerome R Busemeyer. 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, 2006.
- [23] Jan Drugowitsch, Gregory C DeAngelis, Eliana M Klier, Dora E Angelaki, and Alexandre Pouget. Optimal multisensory decision-making in a reaction-time task. *eLife*, 3, jun 2014.
- [24] Jan Drugowitsch, Rubén Moreno-Bote, Anne K Churchland, Michael N Shadlen, and Alexandre Pouget. The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11):3612–3628, 2012.
- [25] Digby Elliott, Steven Hansen, Jocelyn Mendoza, and Luc Tremblay. Learning to optimize speed, accuracy, and energy expenditure: a framework for understanding speed-accuracy relations in goal-directed aiming. *Journal of motor behavior*, 36(3):339–351, 2004.
- [26] William Estes and James Straughan. Analysis of a verbal conditioning situation in terms of statistical learning theory. *Journal of Experimental Psychology*, 47(4):225–234, 1954.
- [27] William K Estes. Toward a statistical theory of learning. *Psychological Review*, 57(2):94, 1950.
- [28] Jean-Claude Falmagne, Stephen Cohen, and A Dwivedi. Two-choice reactions as an ordered memory scanning process. *Attention and Performance V*, pages 296–344, 1975.
- [29] Tim W Fawcett, Benja Fallenstein, Andrew D Higginson, Alasdair I Houston, Dave E W Mallpress, Pete C Trimmer, John M McNamara, and Others. The evolution of decision rules in complex environments. *Trends in Cognitive Sciences*, 18(3):153–161, 2014.

BIBLIOGRAPHY

- [30] Julian Feldman and Joe F Hanna. The structure of responses to a sequence of binary events. *Journal of Mathematical Psychology*, 3(2):371–387, July 1966.
- [31] Julian Feldman, Joe F Hanna, Matt Jones, Winston R Sieck, W K Estes, J H Straughan, Jerome L Myers, Norman H Anderson, Michael C Mozer, Sachiko Kinoshita, Michael Shettel, Harold Pashler, Gordon C Baylis, Donald Laming, Marc Guitart-Masip, Emrah Duzel, Ray Dolan, and Peter Dayan. Effect of first-order conditional probability in a two-choice learning situation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 3(4):180–193, jul 2014.
- [32] Samuel W Fernberger. Interdependence of judgments within the series for the method of constant stimuli. *Journal of Experimental Psychology*, 3(2):126, 1920.
- [33] Merrill M Flood. Environmental non-stationarity in a sequential decision-making experiment. In *Decision Processes*, pages 287–299. Wiley, 1954.
- [34] Ingo Fründ, Felix A Wichmann, and Jakob H Macke. Quantifying the effect of intertrial dependence on perceptual decisions. *Journal of Vision*, 14(7), jun 2014.
- [35] Juan Gao, KongFatt Wong-Lin, Philip Holmes, Patrick Simen, and Jonathan D Cohen. Sequential effects in two-choice reaction time tasks: decomposition and synthesis of mechanisms. *Neural Computation*, 21(9):2407–2436, 2009.
- [36] Crispin Gardiner. *Stochastic methods*, volume 4. Springer Berlin, 2009.
- [37] Christopher M Glaze, Alexandre L S Filipowicz, Joseph W Kable, Vijay Balasubramanian, and Joshua I Gold. A bias-variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*, 2(3):213, 2018.
- [38] Joshua I Gold and Michael N Shadlen. Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36(2):299–308, 2002.
- [39] Joshua I Gold and Michael N Shadlen. The neural basis of decision making. *Annu. Rev. Neurosci.*, 30:535–574, 2007.
- [40] Stephanie Goldfarb, KongFatt Wong-Lin, Michael Schwemmer, Naomi Ehrich Leonard, and Philip Holmes. Can post-error dynamics explain sequential reaction time patterns? *Frontiers in Psychology*, 3:213, 2012.

BIBLIOGRAPHY

- [41] Deborah M Gordon. Behavioral flexibility and the foraging ecology of seed-eating ants. *The American Naturalist*, 138(2):379–411, 1991.
- [42] Ming Hsu, Meghana Bhatt, Ralph Adolphs, Daniel Tranel, and Colin F Camerer. Neural systems responding to degrees of uncertainty in human decision-making. *Science*, 310(5754):1680–1683, 2005.
- [43] Eun Jung Hwang, Jeffrey E. Dahlen, Madan Mukundan, and Takaki Komiyama. History-based action selection bias in posterior parietal cortex. *Nature Communications*, 8(1):1–14, 2017.
- [44] Murray E Jarvik. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology*, 41(4):291, 1951.
- [45] Matt Jones, Tim Curran, Michael C Mozer, and Matthew H Wilder. Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, 120(3):628, 2013.
- [46] Matt Jones, Sachiko Kinoshita, and Michael C Mozer. Optimal response initiation: Why recent experience matters. In *Advances in Neural Information Processing Systems*, pages 785–792, 2009.
- [47] Matt Jones and Winston R Sieck. Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4):626–640, 2003.
- [48] Zachary P Kilpatrick. Synaptic mechanisms of interference in working memory. *Scientific Reports*, 8(1):7879, 2018.
- [49] Timothy Doyeon Kim, Mohammad Kabir, and Joshua I Gold. Coupled decision processes update and maintain saccadic priors in a dynamic environment. *Journal of Neuroscience*, 37(13):3632–3645, 2017.
- [50] Neil H Kirby. Sequential effects in two-choice reaction time: Automatic facilitation or subjective expectancy? *Journal of Experimental Psychology: Human Perception and Performance*, 2(4):567, 1976.
- [51] Donald Laming. Choice reaction performance following an error. *Acta Psychologica*, 43(3):199–224, 1979.
- [52] Kenneth W Latimer, Jacob L Yates, Miriam L R Meister, Alexander C Huk, and Jonathan W Pillow. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, 349(6244):184–187, 2015.

BIBLIOGRAPHY

- [53] David A Lieberman, A Carina, M Vogel, and Jaime Nisbet. Why do the effects of delaying reinforcement in animals and delaying feedback in humans differ? A working-memory analysis. *Quarterly Journal of Experimental Psychology*, 61(2):194–202, 2008.
- [54] R Duncan Luce. *Response times: Their role in inferring elementary mental organization*. Number 8. Oxford University Press on Demand, 1986.
- [55] Nicholas H Lurie and Jayashankar M Swaminathan. Is timely information always better? The effect of feedback frequency on decision making. *Organizational Behavior and Human Decision Processes*, 108(2):315–329, 2009.
- [56] Gaurav Malhotra, David S Leslie, Casimir J H Ludwig, and Rafal Bogacz. Time-varying decision boundaries: insights from optimality analysis. *Psychonomic Bulletin & Review*, 25(3):971–996, 2018.
- [57] Helen Susannah Moat, Chester Curme, Adam Avakian, Dror Y Kenett, H Eugene Stanley, and Tobias Preis. Quantifying wikipedia usage patterns before stock market moves. *Scientific Reports*, 3:1801, 2013.
- [58] Michael C Mozer, Sachiko Kinoshita, and Michael Shettel. Sequential dependencies in human behavior offer insights into cognitive control. In *Integrated Models of Cognitive Systems*, pages 180–193. Oxford University Press, may 2007.
- [59] John D Murray, Jorge Jaramillo, and Xiao-Jing Wang. Working memory and decision making in a fronto-parietal circuit model. *Journal of Neuroscience*, pages 317–343, 2017.
- [60] Daniel Navarro and Ben Newell. Information versus reward in a changing world. In *36th Annual Meeting of the Cognitive Science Society*, pages 1054–1059. 2014.
- [61] Khanh P Nguyen, Krešimir Josić, and Zachary P Kilpatrick. Optimizing sequential decisions in the drift-diffusion model. *Journal of Mathematical Psychology*, 2019.
- [62] Ramon Nogueira, Juan M Abolafia, Jan Drugowitsch, Emili Balaguer-Ballester, Maria V Sanchez-Vives, and Rubén Moreno-Bote. Lateral orbitofrontal cortex anticipates choices and integrates prior with current information. *Nature Communications*, 8:14823, 2017.

BIBLIOGRAPHY

- [63] Farzaneh Olianeshad, Maryam Tohidi-Moghaddam, Sajjad Zabbah, and Reza Ebrahimpour. Residual information of previous decision affects evidence accumulation in current decision. *arXiv.org*, nov 2016.
- [64] Charalampos Papadimitriou, Afreen Ferdoash, and Lawrence H Snyder. Ghosts in the machine: memory interference from the previous trial. *Journal of Neurophysiology*, 113(2):567–577, 2015.
- [65] Harold Pashler and Gordon C Baylis. Procedural learning: II. intertrial repetition effects in speeded-choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1):33, 1991.
- [66] Li Qu and Philip David Zelazo. The facilitative effect of positive stimuli on 3-year-olds’ flexible rule use. *Cognitive Development*, 22(4):456–473, 2007.
- [67] Patrick Rabbitt and Subhash Vyas. An elementary preliminary taxonomy for some errors in laboratory choice RT tasks. *Acta Psychologica*, 33:56–76, 1970.
- [68] Patrick Rabbitt and Subhash Vyas. Processing a display even after you make a response to it. How perceptual errors can be corrected. *The Quarterly Journal of Experimental Psychology Section A*, 33(3):223–239, 1981.
- [69] Adrian E Radillo, Alan Veliz-Cuba, Krešimir Josić, and Zachary P Kilpatrick. Evidence accumulation and change rate inference in dynamic environments. *Neural Computation*, 29(6):1561–1610, 2017.
- [70] David Raposo, John P Sheppard, Paul R Schrater, and Anne K Churchland. Multisensory decision-making in rats and humans. *Journal of Neuroscience*, 32(11):3726–3735, 2012.
- [71] Roger Ratcliff. A theory of memory retrieval. *Psychological Review*, 85(2):59, 1978.
- [72] Roger Ratcliff. Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, 92(2):212, 1985.
- [73] Roger Ratcliff and Gail McKoon. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922, 2008.
- [74] Roger Ratcliff and Philip L Smith. A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2):333, 2004.

BIBLIOGRAPHY

- [75] Chris R Reid, Simon Garnier, Madeleine Beekman, and Tanya Latty. Information integration and multiattribute decision making in non-neuronal organisms. *Animal Behaviour*, 100:44–50, 2015.
- [76] Robert J Remington. Analysis of sequential effects on choice reaction times. *Journal of Experimental Psychology*, 82(2):250, 1969.
- [77] Arbora Resulaj, Roozbeh Kiani, Daniel M Wolpert, and Michael N Shadlen. Changes of mind in decision-making. *Nature*, 461(7261):263, 2009.
- [78] Jeffrey D Schall. Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, 2(1):33, 2001.
- [79] Verena D Schmittmann, Ingmar Visser, and Maartje E J Raijmakers. Multiple learning modes in the development of performance on a rule-based category-learning task. *Neuropsychologia*, 44(11):2079–2091, 2006.
- [80] Wolfram Schultz. Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80(1):1–27, 1998.
- [81] Wolfram Schultz, Peter Dayan, and P Read Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.
- [82] Thomas D Seeley, Scott Camazine, and James Sneyd. Collective decision-making in honey bees: how colonies choose among nectar sources. *Behavioral Ecology and Sociobiology*, 28(4):277–290, 1991.
- [83] Michael N Shadlen and William T Newsome. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4):1916–1936, 2001.
- [84] Eric Soetens, Louis Boer, and Johan Hueting. Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 11(5):598–616, 1985.
- [85] Paul Anthony Stevenson and Jan Rillich. The decision to fight or flee—insights into underlying mechanism in crickets. *Frontiers in Neuroscience*, 6:118, 2012.
- [86] Mervyn Stone. Models for choice-reaction time. *Psychometrika*, 25(3):251–260, 1960.
- [87] Richard Sutton and Andrew Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2018.

BIBLIOGRAPHY

- [88] Peter M Todd and Gerd Gigerenzer. Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, 16(3):167–171, 2007.
- [89] Joji Tsunada, Yale Cohen, and Joshua I Gold. Post-decision processing in primate prefrontal cortex influences subsequent choices on an auditory decision-making task. *eLife*, 8:1–21, 2019.
- [90] Anne E Urai, Jan Willem de Gee, Konstantinos Tsetsos, and Tobias H Donner. Choice history biases subsequent evidence accumulation. *eLife*, 8, jul 2019.
- [91] Marius Usher and James L McClelland. The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, 108(3):550, 2001.
- [92] Anna C K Van Duijvenvoorde, Kiki Zanolie, Serge A R B Rombouts, Maartje E J Raijmakers, and Eveline A Crone. Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *Journal of Neuroscience*, 28(38):9495–9503, 2008.
- [93] Alan Veliz-Cuba, Zachary P Kilpatrick, and Krešimir Josić. Stochastic models of evidence accumulation in changing environments. *SIAM Review*, 58(2):264–289, 2016.
- [94] Pascale Waelti, Anthony Dickinson, and Wolfram Schultz. Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, 412(6842):43–48, 2001.
- [95] Abraham Wald and Jacob Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, pages 326–339, 1948.
- [96] Roy A Wise. Dopamine, learning and motivation. *Nature Reviews Neuroscience*, 5(6):483–494, jun 2004.
- [97] Angela J Yu and Jonathan D Cohen. Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems*, 21:1873–1880, 2008.