



University of Pennsylvania  
**ScholarlyCommons**

---

Publicly Accessible Penn Dissertations


---

2019

## The Neural Computations In The Caudate Nucleus For Reward-Biased Perceptual Decision-Making

Yunshu Fan  
*University of Pennsylvania*

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

 Part of the [Neuroscience and Neurobiology Commons](#)

---

### Recommended Citation

Fan, Yunshu, "The Neural Computations In The Caudate Nucleus For Reward-Biased Perceptual Decision-Making" (2019). *Publicly Accessible Penn Dissertations*. 3552.  
<https://repository.upenn.edu/edissertations/3552>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/3552>  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# The Neural Computations In The Caudate Nucleus For Reward-Biased Perceptual Decision-Making

## Abstract

Decision-making is a complex process in which our brain has to combine different sources of information, such as noisy sensory evidence and expected reward, in appropriate ways to obtain the outcome that satisfies the decision-maker. Despite various studies on perceptual decision-making and value-based decision making, it is still unclear how the brain combines sensory and reward information to make a complex decision. A prime candidate for mediating this process is the basal ganglia pathway. This pathway is known to make separate contributions to perceptual decisions based on the interpretation of uncertain sensory evidence and value-based decisions that select among outcome options. To begin to investigate what computations are performed by the brain, particularly in the basal ganglia, we trained monkeys to perform a reward-biased visual motion direction discrimination task and performed single-unit extracellular recordings in the caudate nucleus, the input station in the basal ganglia. Fitting the monkeys' behaviors to a drift-diffusion model, we found that the monkeys used a rational heuristic to combine sensory and reward information. This heuristic is suboptimal but leads to good-enough outcomes. We also found that the monkeys' reward biases were sensitive to the changes in the reward functions from session to session. This adaptive adjustment could be a possible reason underlying the individual variability in their decision strategies. By recording in the caudate nucleus, we found that it is involved in both the decision-formation and evaluation: before the monkey started accumulating sensory evidence, the caudate neurons represented the reward context that could be used to form a reward bias; during decision-formation, some caudate neurons jointly represented sensory evidence and reward information, which could facilitate the combining of sensory and reward information appropriately. After a decision is made, caudate nucleus represented both decision confidence and reward expectation, two evaluation-related quantities that influence the monkeys' subsequent decision behaviors.

## Degree Type

Dissertation

## Degree Name

Doctor of Philosophy (PhD)

## Graduate Group

Neuroscience

## First Advisor

Long . Ding

## Second Advisor

Joshua I. Gold

## Keywords

caudate nucleus, computation, decision-making, drift-diffusion model, optimality, reward

## Subject Categories

Neuroscience and Neurobiology

**THE NEURAL COMPUTATIONS IN THE CAUDATE NUCLEUS FOR REWARD-  
BIASED PERCEPTUAL DECISION-MAKING**

Yunshu Fan

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

**Supervisor of Dissertation**

**Co-Supervisor of Dissertation**

\_\_\_\_\_

Long Ding, Ph.D.  
Research Assistant Professor of Neuroscience

\_\_\_\_\_

Joshua I. Gold, Ph.D.  
Professor of Neuroscience

**Graduate Group Chairperson**

\_\_\_\_\_

Joshua I. Gold, Ph.D., Professor of Neuroscience

**Dissertation Committee**

Nicole C. Rust, Ph.D., Associate Professor of Psychology

Joseph W. Kable, Ph.D., Baird Term Professor of Psychology

Johannes Burge, Ph.D., Associate Professor of Psychology

Bruno B. Averbeck, Ph.D., Section Chief on Learning and Decision Making, National Institute of Health

THE NEURAL COMPUTATIONS IN THE CAUDATE NUCLEUS FOR REWARD-  
BIASED PERCEPTUAL DECISION-MAKING

COPYRIGHT

2019

Yunshu Fan

This work is licensed under the  
Creative Commons Attribution-  
NonCommercial-ShareAlike 4.0  
License

To view a copy of this license, visit

<https://creativecommons.org/licenses/by-nc-sa/4.0/us/>

*Dedicated to Grandpa who showed me how to love, care and support unconditionally,  
and to Bunny whom I believe is sent by Grandpa.*

## ACKNOWLEDGMENT

Seven years ago, when I was going to decide which lab to join as my thesis lab, a senior NGG student told me to choose the lab that I feel like home. So I decided to join Long and Josh's labs and switched to the Neuroscience Graduate Group. Till this day, the feeling of home has never left me because I have always been surrounded by amazing people who helped and supported me along the way.

I was fortunate to have Long and Josh as my advisors and mentors. They not only taught me everything about neuroscience and programming from scratch, but also showed me how to do rigorous science and how to be good mentors and persons who truly care about well-being of others. They lead by example, and I hope I could be like them in the future.

I am also fortunate to have supportive lab mates. Members of the Ding and Gold labs, past and current, have all helped me in various ways, from experiments and analyses to friendship, compassion and encouragement. Special thanks to Taka, also my research collaborator, who is always ready to listen, help and empathize; to Javier, who cheered me up during many of my darkest moments; to Yin, who gave me so much good advice; to Lalitta, a true friend; to Kyra, whom I can share my interests in crafts with; to Alice, who is always cheering me up; to Matt, who consistently offered help even though we did not overlap; to Alex, for your positivity; to Adrian, who showed me to be the change you want to see. I also want to thank the Computational Neuroscience Initiative, a community that shares not only knowledge, but also the passion for learning. Over the years, I also met people who went through tremendous setbacks in research

and life. Their resilience has been my source of courage. I don't want to reveal their names, but I want to thank them for being my inspirations.

I'm fortunate to have met mentors outside lab who gave me timely advice and encouragement. Thank you to Dr. Nicole Rust, the chair of my committee, for noticing that I was driven by fear and encouraging me to chase my passion. Thank you to Dr. Gregory Horwitz, for your encouraging words. Thank you Dr. Adam Resnick—you saw the potential in me and you helped me manifest it.

I want to thank Morgan Taylor, Madineh Sarvestani, Evelyn Tang, Seha Kim, Noam Roth, Shachee Doshi, Sarah Ly, Farzaneh Najafi and AWIS mentor circle for being my peer mentor and give me support and guidance.

I want to thank my two dance teacher, Ms. JoAnna Turner and A.T. Moffet. You believed that I can learn something completely new, and you showed me how to be an inspiring teacher.

Being an international student is not easy. I want to thank those who made this foreign country more like home to me: the lammarino Family, Nancy Yu and Sally Zhou.

Thank you to my old friends Jiang Dan, Zhang Ketian, Shi Xu, Chen Chong, Guo Xuanzhong, Wang Chen, You Xuedi, Kang Jian, Liu Yang, Liu Yuejiang, Liu Chen, Yang Aixuan, and Zhang Honglin (who taught me how to read Matlab code, like what does “;” mean, in front of the ENIAC). We came to study abroad around the same time. We witnessed each other's joy, sorrow and growth. I'm glad we still check in with each other.

Thank you to the friends I met in graduate school. I have learned so much by being part of the NGG, such a warm and supportive community. Thank you, Jingwen, Yueyao, Tianxin, Nils and Tina, Julia, Kim, Anna Li-Stern, Felicia, Patti, and Meilin. You

taught me that I don't have to be afraid of going into a new environment, and it is possible to make new true friends.

Also thank you to my friends back in China, Nie Jingyi, Li Meng, Liu Feng, Zhao Xin, Li Wei, Song Yingda, Zhang Juan. You showed me that distance does not matter if we truly care about each other.

Finally, I would not have been able to make this far without the love and support from my family. Mom and Grandma will continue to be my inspiration, and I know Grandpapa will keep me company. Meanwhile, I'm also grateful for meeting people who treated me as a part of their family, with their unconditional love and timely help. Therefore, I want to give my special thanks also to my extended family—Bihui and John Melidosian, John and Carmel Iammarino, the Joshi family, and the treasure I found in grad school—Sidd.



# ABSTRACT

## THE NEURAL COMPUTATIONS IN THE CAUDATE NUCLEUS FOR REWARD-BIASED PERCEPTUAL DECISION-MAKING

Yunshu Fan

Long Ding and Joshua I. Gold

Decision-making is a complex process in which our brain has to combine different sources of information, such as noisy sensory evidence and expected reward, in appropriate ways to obtain the outcome that satisfies the decision-maker. Despite various studies on perceptual decision-making and value-based decision making, it is still unclear how the brain combines sensory and reward information to make a complex decision. A prime candidate for mediating this process is the basal ganglia pathway. This pathway is known to make separate contributions to perceptual decisions based on the interpretation of uncertain sensory evidence and value-based decisions that select among outcome options. To begin to investigate what computations are performed by the brain, particularly in the basal ganglia, we trained monkeys to perform a reward-biased visual motion direction discrimination task and performed single-unit extracellular recordings in the caudate nucleus, the input station in the basal ganglia. Fitting the monkeys' behaviors to a drift-diffusion model, we found that the monkeys used a rational heuristic to combine sensory and reward information. This heuristic is suboptimal but leads to good-enough outcomes. We also found that the monkeys' reward biases were sensitive to the changes in the reward functions from session to session. This adaptive adjustment could be a possible reason underlying the individual variability in their decision strategies. By recording in the caudate nucleus, we found that it is involved in

both the decision-formation and evaluation: before the monkey started accumulating sensory evidence, the caudate neurons represented the reward context that could be used to form a reward bias; during decision-formation, some caudate neurons jointly represented sensory evidence and reward information, which could facilitate the combining of sensory and reward information appropriately. After a decision is made, caudate nucleus represented both decision confidence and reward expectation, two evaluation-related quantities that influence the monkeys' subsequent decision behaviors.

# TABLE OF CONTENTS

<b>THE NEURAL COMPUTATIONS IN THE CAUDATE NUCLEUS FOR REWARD-BIASED PERCEPTUAL DECISION-MAKING.....</b>	<b>II</b>
<b>ACKNOWLEDGMENT .....</b>	<b>IV</b>
<b>ABSTRACT.....</b>	<b>VII</b>
<b>LIST OF TABLES .....</b>	<b>XII</b>
<b>LIST OF ILLUSTRATIONS .....</b>	<b>XIII</b>
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
<b>Computational framework for perceptual decision-making.....</b>	<b>2</b>
<b>Computational framework for value-based decision-making .....</b>	<b>7</b>
<b>A possible computational framework that combines perceptual decision-making and     value-based decision making .....</b>	<b>8</b>
<b>Caudate nucleus and decision-making .....</b>	<b>10</b>
<b>Reference.....</b>	<b>13</b>

<b>CHAPTER 2: ONGOING, RATIONAL CALIBRATION OF REWARD-DRIVEN PERCEPTUAL BIASES .....</b>	<b>18</b>
Introduction .....	19
Results .....	22
Discussion .....	51
 <b>CHAPTER 3: NEURAL REPRESENTATION OF SENSORY AND REWARD INFORMATION IN THE CAUDATE NUCLEUS IN REWARD-BIASED PERCEPTUAL DECISION-MAKING .....</b>	<b>80</b>
Introduction .....	80
Results .....	81
Discussion .....	97
Materials and methods .....	100
Reference .....	105
 <b>CHAPTER 4: CONFIDENCE AND REWARD EXPECTATION ARE REPRESENTED IN CAUDATE POST-DECISION ACTIVITY .....</b>	<b>110</b>
Introduction .....	110
Results .....	112

<b>Discussion .....</b>	<b>124</b>
<b>Materials and Methods .....</b>	<b>128</b>
<b>Reference .....</b>	<b>137</b>
<b>CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS .....</b>	<b>140</b>
<b>Experimental/Task design.....</b>	<b>141</b>
<b>Caudate nucleus and reward-biased perceptual decision-making .....</b>	<b>145</b>
<b>New theoretical frameworks .....</b>	<b>148</b>
<b>Towards computational psychiatry.....</b>	<b>149</b>
<b>Reference .....</b>	<b>151</b>

## LIST OF TABLES

Table 2.1 Best fitting DDM parameters. ....	31
Table 2.2 Model comparisons. ....	32
Table 4.1. Distribution of confidence- and reward expectation-representing neurons in sessions with confidence-related sequential effects, reward expectation-related sequential effects and no sequential effects. ....	124

## LIST OF ILLUSTRATIONS

Figure 2.1. Theoretical framework and task design. ....	21
Figure 2.2. Relationships between sensitivity and bias from logistic fits to choice data. .....	25
Figure 2.2-figure supplement 1. Monkeys showed minimal sequential choice biases. ....	25
Figure 2.2-figure supplement 2. The optimal bias decreases with increasing sensitivity. .....	26
Figure 2.3. Comparison of choice and RT data to HDDM fits with both momentary- evidence ( <i>me</i> ) and decision-rule ( <i>z</i> ) biases. ....	29
Figure 2.3–figure supplement 1. Qualitative comparison between the monkeys' RT distribution and DDM predictions. ....	28
Figure 2.3–figure supplement 2. Fits to a DDM with collapsing bounds. ....	30
Figure 2.4. Actual versus optimal adjustments of momentary-evidence ( <i>me</i> ) and decision-rule ( <i>z</i> ) biases. ....	34
Figure 2.4–figure supplement 1. Estimates of momentary-evidence ( <i>me</i> ) and decision- rule ( <i>z</i> ) biases using the collapsing-bound DDM fits. ....	35
Figure 2.5. Predicted versus optimal reward per trial (RTrial). ....	37
Figure 2.5–figure supplement 1. Predicted versus optimal reward rate (RR). ....	38
Figure 2.6. Relationships between adjustments of momentary-evidence ( <i>me</i> ) and decision-rule ( <i>z</i> ) biases and RTrial function properties. ....	39
Figure 2.6–figure supplement 1. The monkeys' momentary-evidence ( <i>me</i> ) and decision- rule ( <i>z</i> ) adjustments reflected RR function properties. ....	41

Figure 2.6–figure supplement 2: The HDDM model fitting procedure does not introduce spurious correlations between patch orientation and <i>me</i> value. ....	42
Figure 2.6–figure supplement 3. The correlation between fitted and conditionally optimal adjustments was stronger for the real, session-by-session data than for unmatched sessions. ....	43
Figure 2.7. Relationships between starting and ending values of the satisficing, reward function gradient-based updating process. ....	45
Figure 2.7–figure supplement 1. RR gradient trajectories color-coded by the end points of the <i>me/z</i> patterns. ....	46
Figure 2.8. The satisficing reward function gradient-based model. ....	47
Figure 2.8–figure supplement 1. Predictions of a RR gradient-based model. ....	50
Figure 2.8–figure supplement 2. Dependence of the orientation and area of the near-optimal RTrial patch on parameters reflecting internal decision process and external task specifications. ....	56
Figure 2.8–figure supplement 3: The joint effect of DDM model parameters <i>a</i> (governing the speed-accuracy trade-off) and <i>k</i> (governing perceptual sensitivity) on the shape of the reward function. ....	58
Figure 2.8–figure supplement 4. Effects of the shape of the reward function on deviations from optimality. ....	61
Figure 2.8–figure supplement 5. Hypothetical neural activity encoding a reward-biased perceptual decision variable. ....	63
Figure 3.1 Monkeys showed biases toward choices associated with large reward. ....	82
Figure 3.2 Reward context representations before visual stimulus onset. ....	84



Figure 3.2-figure supplement 1. Diverse temporal dynamics of reward context representations before visual stimulus onset. ....	86
Figure 3.3 Caudate activity reflected motion strength, reward context, choice, and the expected reward size associated with the choice. ....	89
Figure 3.3-figure supplement 1. Example neurons with reward context modulation. ....	90
Figure 3.3-figure supplement 2: Example neurons with different kinds of task-relevant modulations. ....	92
Figure 3.3-figure supplement 3: Comparison between decision variable and caudate neural activity. ....	94
Figure 3.4 Reaction time (RT) is represented in caudate late-decision and post-decision activities. ....	96
Figure 4.1. Confidence and reward expectation depended on decision time, motion coherence, and reward asymmetry-induced biases. ....	114
Figure 4.2 Confidence- and reward expectation-related sequential influence on monkeys' choice and RT. ....	117
Figure 4.2-figure supplement 1. ....	119
Figure 4.3 Example post-decision caudate neural activities that resemble confidence and reward expectation. ....	121
Figure 4.4 Confidence and reward expectation correlate with the post-saccade activity in subpopulation of caudate neurons. ....	123
Figure 4.5 Related to “computing confidence” in Methods: Computing the probability of making a rightward choice at time T for a given motion coherence. ....	124

## Chapter 1: Introduction

Yunshu Fan, Joshua I Gold, Long Ding

Decision-making often requires combining evidence for and against different options and their expected outcomes. For example, when we decide “should I eat more chocolate”, and encounter a claim that “chocolate is healthy”, the decision could be influenced by whether the claim is from a peer-reviewed research article or tabloid, and by the desired outcome: “I hope it is true because I love chocolate!” Similarly, when we decide whether to keep staying in academia, and encountered advice that “a faculty position is harder to get nowadays”, the decision could be influenced by whether the advice is written based on nation-wide statistical studies of faculty applicants or on anecdotes from postdocs who failed to get faculty positions several times in a row, and by an internal preference: “I really like doing research.” The ability of our brain to perform computations that collect and interpret evidence with various levels of reliability and combine that with our internal preference for specific outcomes gives us the capacity to make complex decisions.

In this introduction, I will begin by reviewing computational frameworks used for studying decision-making driven by sensory evidence (perceptual decision-making) and decision-making driven by outcomes (value-based decision-making) and how the two could be combined. I will then turn to the caudate nucleus in the basal ganglia, a brain region that may play a key role in combining sensory evidence and reward outcomes in decision-making, with a focus on the anatomical and neurophysiological findings that support this role.

## Computational framework for perceptual decision-making

A perceptual decision is a categorical judgment about the state of the environment based on the noisy data provided by the sensory system. We usually make such decisions without even realizing it. For example, before crossing a road, we judge whether a car is approaching us; on a trail, we might judge the wind direction based on the fluttering leaves on a tree; an experienced cook might decide whether a steak is cooked based on the sizzling sound; when tuning a guitar, we decide whether the pitch is higher or lower than the standard; a hungry kid might know whether dinner is ready by sniffing the air. In each case, the sensory inputs, like motion, pitch, odor, etc. that are usually noisy. Therefore, the decision is not a simple reflex, but the result of a deliberative process.

According to the signal detection theory (SDT, Green and Swets, 1966), the perceptual decision-making process could be formalized as a form of statistical inference. The possible alternatives corresponding to the different states of the world could be thought of as hypotheses ( $H$ ), and the sensory input as evidence ( $e$ ). A decision is made by selecting the most probable hypothesis supported by the evidence; i.e., the posterior probability given the sensory input ( $P(H|e)$ ). When there are two

alternatives,  $H_1$  and  $H_2$ , the selection process is equivalent to comparing the ratio of the posteriors. When  $\frac{P(H_1|e)}{P(H_2|e)} > 1$ , it suggests that  $H_1$  is more accurate than  $H_2$ , thus  $H_1$

should be chosen; conversely, when  $\frac{P(H_1|e)}{P(H_2|e)} < 1$ ,  $H_2$  is more likely to be true, therefore

should be chosen. We can define a decision variable (DV) as follows:

$$DV = \frac{P(H_1|e)}{P(H_2|e)} \quad (\text{Eq. 1})$$

According to Bayes Rule,  $(H_i|e) = P(e|H_i) \times \frac{P(H_i)}{P(e)}$ , where  $P(e|H_i)$  is the likelihood of observing the specific evidence if that hypothesis is true. If the evidence supports one alternative, say  $H_1$ , more than the other, its likelihood under  $H_1$  should be much larger than its likelihood under  $H_2$ .  $P(H_i)$  is the preconceived probability of  $H_i$  being true, and  $P(e)$  is the probability of observing the evidence regardless of any particular hypothesis being true. These quantities are also referred to as Priors. The prior over the evidence  $P(e)$  is canceled out when computing the DV as in Eq. 1:

$$DV = \frac{P(H_1|e)}{P(H_2|e)} = \frac{P(e|H_1)}{P(e|H_2)} \times \frac{P(H_1)}{P(H_2)} \quad (\text{Eq. 2})$$

Consequently, in perceptual decision-making, a decision is influenced by the likelihood ratio  $\left(\frac{P(e|H_1)}{P(e|H_2)}\right)$  and the prior ratio  $\left(\frac{P(H_1)}{P(H_2)}\right)$ .

Eq.2 assumes that there is only one piece of evidence for making the decision. If the decision is based on multiple pieces of evidence, and if we assume that each piece of evidence is independent from another, then the likelihood of observing all the evidence would be the same as the product of the likelihood of observing each piece of evidence. Therefore, the likelihood ratio in Eq. 2 can be expanded in the following way:

$$\begin{aligned} \frac{P(e|H_1)}{P(e|H_2)} &= \frac{P(e_1, e_2, \dots, e_n | H_1)}{P(e_1, e_2, \dots, e_n | H_2)} \\ &= \frac{P(e_1|H_1) \times P(e_2|H_1) \times \dots \times P(e_n|H_1)}{P(e_1|H_2) \times P(e_2|H_2) \times \dots \times P(e_n|H_2)} \\ &= \frac{P(e_1|H_1)}{P(e_1|H_2)} \times \frac{P(e_2|H_1)}{P(e_2|H_2)} \times \dots \times \frac{P(e_n|H_1)}{P(e_n|H_2)} \quad (\text{Eq. 3}) \end{aligned}$$

We can take the log on both sides, such that Eq.2 becomes:

$$\log DV = \log \frac{P(H_1|e)}{P(H_2|e)} = \log \frac{P(H_1)}{P(H_2)} + \sum_{i=1}^n \log \frac{P(e_i|H_1)}{P(e_i|H_2)} \quad (\text{Eq. 4})$$

We can now redefine the decision variable as the log of the posterior ratio:

$$DV_{new} = \log DV$$

and compare the new DV with 0 when making a decision

Thus, under the assumption of independence of evidence, Eq. 4 suggests that when there are multiple pieces of evidence, a decision maker can simply add all the log likelihood ratios together.

This formulation also gives us an easy way to deal with each new piece of evidence. If we think of each piece of evidence as the sensory input at a given time (t), accumulating additional piece of evidence at time t+1 is equivalent to update the DV by adding the log likelihood ratio of the new evidence and comparing the updated DV with 0.

This framework is the basic form of the sequential probability ratio test (SPRT, Barnard, 1946; Wald, 1947).

If the decision maker only cares about which hypothesis is more probable than the other so as to select one option, then the magnitude of DV does not matter. However, the magnitude of DV influences accuracy. For example, even though  $DV = -0.1$  and  $DV = -1$  both support choosing  $H_2$ , the first suggests a lower certainty (a.k.a., accuracy) in the choice than the second.

If the sensory evidence indeed supports one choice, i.e., there is signal in the sensory input (not pure noise), then adding additional evidence can, in theory, increase the magnitude of DV, therefore making the decision more accurate. In other words, accumulating noisy evidence can strengthen the signal by averaging out sensory noise.

Therefore, if the decision makers can determine how long to accumulate evidence before committing to a decision, they can control the overall accuracy of the decision by setting the bounds for DV to reach before committing to one decision or another. They can aim for more accuracy by accumulating more evidence, which takes longer time, or he/she can aim for less accurate but faster decision by accumulate less evidence. This balance is known as the “speed-accuracy trade-off” (Palmer et al., 2005; Forstmann et al., 2010; Hanks et al., 2011).

To summarize, the decision-making process can be described as follows: update the DV by accumulating evidence, and compare the DV with the bounds. If the DV reaches a bound, stop accumulating and commit to the decision represented by that bound; if not, accumulate more evidence.

When we treat time as a continuous term instead of discrete, this process is formulated as the drift-diffusion model (DDM), the model that I will use in the following chapters. The DDM was first applied to psychology/neuroscience study to explain the memory retrieval process (Ratcliff, 1978). Since then, it has been used to explain the

behaviors in a variety of decision-making tasks (ref). Its quantitative framework has facilitated the discovery of neural correlates of behavior in many brain areas (ref).

In the DDM, momentary evidence is modeled as a Gaussian distribution, i.e., assuming noise in the sensory input is independent from time to time. The mean of the momentary evidence is a monotonic function of the signal strength of the sensory input. For example, in a motion discrimination task, momentary evidence is typically modeled as the coherence of the moving dots multiplied by a scaling factor (Palmer et al., 2005). This way of modeling the momentary evidence is supported by the finding that motion sensitive neurons in visual cortex that are involved in motion discrimination scaled their responses with coherence (Salzman et al., 1992; Britten et al., 1993).

The DV is the time integral of the momentary evidence and is constantly compared with two bounds that represent the total amount of evidence needed to commit to the two options, respectively. The DV will gradually drift to one bound or the other over time due to the signal in the sensory input. When the sensory evidence is strong, the DV will drift towards and reach a bound faster; when the sensory evidence is weaker, the DV will drift towards and reach a bound slower.

Neural correlates of the momentary evidence should be sensitive to the stimulus strength and not change with time. In contrast, neural correlates of the DV should reflect both the stimulus strength and the evidence accumulation over time. Neural correlates of momentary evidence are often found in sensory areas. For example, visual motion evidence is conveyed by the motion sensitive neurons in extrastriate areas MT and MST (Britten et al., 1992; Celebrini and Newsome, 1994; Britten et al., 1996). Evidence about vibrotactile frequency was found in the primary somatosensory cortex (Mountcastle et al., 1990; Salinas et al., 2000). Neurons in the middle-lateral and anterolateral belt

region of the auditory cortex encode sound frequency evidence in high/low pitch discrimination(Tsunada et al., 2015). Neural correlates of DV have been found in sensorimotor or motor-related areas, such as, the lateral intraparietal area (Roitman and Shadlen, 2002), frontal eye field(Ding and Gold, 2012a), and the pre-motor cortex(Suriya-Arunroj and Gail, 2019).

So far, I have described computational framework dealing with decision-making based on sensory evidence. Next I am going to introduce the computational framework dealing with decision-making based on outcomes from choices (value-based decision-making).

### **Computational framework for value-based decision-making**

Value-based decision-making is a process that is primarily driven by different outcomes. It is usually studied in tasks where the sensory input does not have ambiguity and in the context of economic decisions. A decision is made by comparing the expected utility (EU, a.k.a. value) of different outcomes. The expected utility of an option is computed by multiplying the subjective estimate of the magnitude of the outcome (usually in the form of reward (R)) with the probability of obtaining that outcome (usually set ahead of time and therefore does not need computing):

$$EU_i = R_i \times P(R_i) \quad (Eq. 5)$$

For example, the probability of the rewards might be manipulated as follows: the subject is presented with one option that is rewarded 70% of the time, and the other option 50% of the time. The magnitude of the reward could also be manipulated so that one option appears more favorable than the other. A search for brain regions with neural activity that reflects reward magnitude, reward probability or EU, has demonstrated value



representation in many brain areas, such as the medial prefrontal cortex, orbitofrontal cortex, lateral intraparietal area, caudate nucleus, putamen and ventral striatum (Lauwereyns et al., 2002; Samejima et al., 2005; Ding and Hikosaka, 2006; Nakamura and Hikosaka, 2006; Padoa-Schioppa and Assad, 2006, 2008; Tom et al., 2007; Lau and Glimcher, 2008).

### **A possible computational framework that combines perceptual decision-making and value-based decision making**

Thus far, I have described computational frameworks for sensory-based and value-based decision-making separately. However, as in the example I gave at the very beginning, real-world choices usually involve combining sensory evidence with non-sensory factors, such as the preference driven by outcomes. One way to combine them is incorporating the expected utility theory with the DDM.

If we assume that a decision is made by choosing the option with the larger EU, we can still construct a DV that is the log ratio of the expected utilities of the two options and compare it with 0. The expected utility of each option can be computed based on Eq. 5. We can approximate the probability of obtaining the reward associated with an option ( $P(R_i)$ ) with the posterior of the hypothesis of that option being true ( $P(H_i|e)$ ). In

this way, the new DV becomes:

$$\begin{aligned} DV &= \log \frac{EU_1}{EU_2} \\ &= \log \frac{P(H_1|e) \times R_1}{P(H_2|e) \times R_2} \end{aligned}$$

$$= \log \frac{P(H_1)}{P(H_2)} + \sum_{i=1}^n \log \frac{P(e_i|H_1)}{P(e_i|H_2)} + \log \frac{R_1}{R_2} \quad (\text{Eq. 6})$$

Eq. 6 suggested that the decision-making process combining sensory and non-sensory factors could be regarded as a modified version of evidence accumulation. The subjective difference between the two rewards ( $\log \frac{R_1}{R_2}$ ) could change the DV independent of evidence accumulation. Meanwhile,  $\log \frac{R_1}{R_2}$  could be parsed into individual evidence:

$$\begin{aligned} \log \frac{P(H_1)}{P(H_2)} + \sum_{i=1}^n \log \frac{P(e_i|H_1)}{P(e_i|H_2)} + \log \frac{R_1}{R_2} \\ = \log \frac{P(H_1)}{P(H_2)} + \sum_{i=1}^n \log \frac{P(e_i|H_1) \times \sqrt[n]{R_1}}{P(e_i|H_2) \times \sqrt[n]{R_2}} \quad (\text{Eq. 7}) \end{aligned}$$

Eq. 7 suggests that the difference in rewards could also lead to misinterpretation of the evidence so they all seem to support one of the options more than under neutral condition, therefore changing the momentary evidence.

In the DDM framework, these two changes could be implemented as changes in the starting value of the DV and in the momentary evidence.

When the signal strength in the noisy sensory input is constant, the optimal strategy is to adjust only the starting value of the DV; when the signal strength can be variable, the optimal strategy is to adjust both the starting value and the momentary evidence (Bogacz et al., 2006). When making perceptual decisions that are biased towards the percept associated with the larger payoff, human and animal subjects showed high individual variations in whether they adjusted the starting value, the momentary evidence or both (Voss et al., 2004; Simen et al., 2009; Summerfield and

Koechlin, 2010; Leite and Ratcliff, 2011; Mulder et al., 2012; Goldfarb et al., 2014; Cicmil et al., 2015). It is unknown what drives a subject's particular strategy.

### **Caudate nucleus and decision-making**

The caudate nucleus is one of the input stations of the basal ganglia, a network of interconnected subcortical nuclei. Its anatomical connection and physiological properties suggest that it might be involved in combining sensory evidence and reward information in decision-making.

Anatomically, it receives inputs from brain regions that process sensory information, as well as regions that carry reward-related information. For example, the caudate nucleus receives projections from areas such as the MT, MST, LIP and FEF, which have been shown to be involved in an oculomotor decision task by processing the visual motion information (Maunsell and Van Essen, 1983; Selemon and Goldman-Rakic, 1985, 1988; Newsome et al., 1989; Saint-Cyr et al., 1990; Britten et al., 1992, 1996; Salzman et al., 1992; Yeterian and Pandya, 1995; Shadlen and Newsome, 1996; Kim and Shadlen, 1999; Roitman and Shadlen, 2002; Ditterich et al., 2003; Hanks et al., 2006; Ding and Gold, 2012b). It also receives inputs from brain areas carrying reward- or value-related signals, such as the medial prefrontal cortex, orbitofrontal cortex (Haber et al., 1995; Padoa-Schioppa and Assad, 2006; Kable and Glimcher, 2007). In addition, the dopaminergic neurons in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) project densely to the caudate nucleus. These dopaminergic neurons encode reward-prediction error signals, which could further modulate how sensory and reward information are combined, especially during learning (Lak et al., 2019).

Neurophysiological evidence also supports the caudate nucleus's roles in sensory and reward processing. For sensory processing, during a visual decision task and an auditory decision task, neural activity in the striatum was found to correlate with the strength of the sensory evidence (Ding and Gold, 2010; Seo et al., 2012; Wang et al., 2018; Yartsev et al., 2018). Manipulating caudate activity during decision-making in both tasks biased the animals' decisions, suggesting a causal role of the caudate nucleus in interpreting sensory information. In the post-decision period, caudate neural activity was found to correlate with some aspects of the sensory information in task, which could be used as decision monitoring and evaluation (Ding and Gold, 2010; Yanike and Ferrera, 2014). It is worth noting that, in these studies, the reward was identical for both choices. Therefore, their results cannot inform us how reward information and sensory information are combined.

For reward processing, when monkeys were asked to choose from two options with different magnitudes of reward, caudate neurons were found to encode the values of the options during decision and the value of the option chosen after decision (Lau and Glimcher, 2008). In another experimental paradigm, monkeys were trained to make a saccadic eye movement to a target flashed at one of two possible locations, with one location associated with large reward, the other small (or one with reward, the other without) (Kawagoe et al., 1998; Lauwereyns et al., 2002; Ding and Hikosaka, 2006). The monkeys' reaction time was found to be faster towards the large reward target. Neural activity in the caudate nucleus was found to represent the reward-location association. Manipulating the neural activity via dopamine antagonists influences the reward-dependent reaction time (Nakamura and Hikosaka, 2006). In human fMRI studies, caudate BOLD signal was found to represent a bias toward the option with higher reward

probability (Forstmann et al., 2010; Mulder et al., 2012). However, in those studies with reward manipulation, the sensory information was either with 100% certainty (visually instructed) or with a constant level of ambiguity, so their findings also cannot address how sensory and reward information are combined in the caudate nucleus.

My thesis examines how the brain combines sensory and reward information during decision-making and the computational roles of the caudate nucleus before, during and after such decisions. To this aim, I trained monkeys to perform a reward-biased perceptual decision-making task and recorded in their caudate nucleus while they were performing the task. In the first chapter I will present my findings on the strategies used by individual monkeys to combine sensory and reward information and the common principles underlying their strategies; in the second chapter, I will describe the neural representation of information in the caudate nucleus with a focus on how it contributes to combining sensory and reward information before, during and after the decision-making; in the third chapter, I will focus on the evaluative nature of the post-decision activity in the caudate nucleus.

## Reference

- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113:700–765.
- Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA (1996) A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci* 13:87–100.
- Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12:4745–4765.
- Britten KH, Shadlen MN, Newsome WT, Movshon JA (1993) Responses of neurons in macaque MT to stochastic motion signals. *Vis Neurosci* 10:1157–1169 .
- Celebrini S, Newsome WT (1994) Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *J Neurosci* 14:4109–4124 .
- Cicmil N, Cumming BG, Parker AJ, Krug K (2015) Reward modulates the effect of visual cortical microstimulation on perceptual decisions. *eLife* 4:e07832.
- Ding L, Gold JI (2010) Caudate encodes multiple computations for perceptual decisions. *J Neurosci* 30:15747–15759.
- Ding L, Gold JI (2012) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cereb Cortex* 22:1052–1067.
- Ding L, Hikosaka O (2006) Comparison of reward modulation in the frontal eye field and caudate of the macaque. *J Neurosci* 26:6695–6703.

Ditterich J, Mazurek ME, Shadlen MN (2003) Microstimulation of visual cortex affects the speed of perceptual decisions. *Nat Neurosci* 6:891–898.

Forstmann BU, Brown S, Dutilh G, Neumann J, Wagenmakers E-J (2010) The neural substrate of prior information in perceptual decision making: a model-based analysis. *Front Hum Neurosci* 4:40.

Goldfarb S, Leonard NE, Simen P, Caicedo-Núñez CH, Holmes P (2014) A comparative study of drift diffusion and linear ballistic accumulator models in a reward maximization perceptual choice task. *Front Neurosci* 8:148.

Haber SN, Kunishio K, Mizobuchi M, Lynd-Balta E (1995) The orbital and medial prefrontal circuit through the primate basal ganglia. *J Neurosci* 15:4851–4867.

Hanks TD, Ditterich J, Shadlen MN (2006) Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nat Neurosci* 9:682–689.

Hanks TD, Mazurek ME, Kiani R, Hopp E, Shadlen MN (2011) Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J Neurosci* 31:6339–6352 .

Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. *Nat Neurosci* 10:1625–1633.

Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1:411–416.

Kim J-N, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat Neurosci* 2:176–185.

Lak A, Okun M, Moss M, Gurnani H, Farrell K, Wells MJ, Reddy CB, Kepecs A, Harris KD, Carandini M (2019) Neural basis of learning guided by sensory confidence and reward value. *bioRxiv*:411413.

- Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58:451–463.
- Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002) A neural correlate of response bias in monkey caudate nucleus. *Nature* 418:413–417.
- Leite FP, Ratcliff R (2011) What cognitive processes drive response biases? A diffusion model analysis. *Judgm Decis Mak* 6:651–687.
- Maunsell JHR, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol* 49:1127-1147.
- Mountcastle VB, Steinmetz MA, Romo R (1990) Frequency discrimination in the sense of flutter: Psychophysical measurements correlated with postcentral events in behaving monkeys. *J Neurosci* 10:3032-3044.
- Mulder MJ, Wagenmakers E-J, Ratcliff R, Boekel W, Forstmann BU (2012) Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J Neurosci* 32:2335–2343.
- Nakamura K, Hikosaka O (2006) Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *J Neurosci* 26:5360–5369.
- Newsome WT, Britten KH, Movshon JA (1989) Neuronal correlates of a perceptual decision. *Nature* 341:52–54.
- Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441:223–226.
- Padoa-Schioppa C, Assad JA (2008) The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat Neurosci* 11:95–102.
- Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and



- accuracy of a perceptual decision. *J Vis* 5:376–404.
- Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22:9475–9489.
- Saint-Cyr JA, Ungerleider LG, Desimone R (1990) Organization of visual cortical inputs to the striatum and subsequent outputs to the pallido-nigral complex in the monkey. *J Comp Neurol* 298: 129-156.
- Salinas E, Hernández A, Zainos A, Romo R (2000) Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. *J Neurosci* 20:5503-5515.
- Salzman CD, Murasugi CM, Britten KH, Newsome WT (1992) Microstimulation in visual area MT: effects on direction discrimination performance. *J Neurosci* 12:2331–2355.
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340.
- Selemon LD, Goldman-Rakic PS (1985) Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. *J Neurosci* 5:776–794.
- Selemon LD, Goldman-Rakic PS (1988) Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *J Neurosci* 8:4049-4068.
- Seo M, Lee E, Averbeck BB (2012) Action Selection and Action Value in Frontal-Striatal Circuits. *Neuron* 74:947–960.
- Shadlen MN, Newsome WT (1996) Motion perception: seeing and deciding. *Proc Natl*

Acad Sci U S A 93:628–633.

Simen P, Contreras D, Buck C, Hu P, Holmes P, Cohen JD (2009) Reward Rate

Optimization in Two-Alternative Decision Making: Empirical Tests of Theoretical Predictions. *J Exp Psychol Hum Percept Perform* 35:1865–1897.

Summerfield C, Koechlin E (2010) Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. *Front Hum Neurosci* 4:208.

Suriya-Arunroj L, Gail A (2019) Complementary encoding of priors in monkey frontoparietal network supports a dual process of decision-making. *Elife* 8:1–21.

Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The neural basis of loss aversion in decision-making under risk. *Science* 315:515-518.

Tsunada J, Liu ASK, Gold JI, Cohen YE (2015) Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nat Neurosci* 19:135–142.

Voss A, Rothermund K, Voss J (2004) Interpreting the parameters of the diffusion model: an empirical validation. *Mem Cognit* 32:1206–1220.

Wang L, Rangarajan K V., Gerfen CR, Krauzlis RJ (2018) Activation of Striatal Neurons Causes a Perceptual Decision Bias during Visual Change Detection in Mice. *Neuron* 97:1369-1381.e5.

Yanike M, Ferrera VP (2014) Interpretive monitoring in the caudate nucleus. *Elife* 3:1–16.

Yartsev MM, Hanks TD, Yoon AM, Brody CD (2018) Causal contribution and dynamical encoding in the striatum during evidence accumulation. *Elife* 7:1–24.

Yeterian EH, Pandya DN (1995) Corticostriatal connections of extrastriate visual areas in rhesus monkeys. *J Comp Neurol* 352:436–457.

## CHAPTER 2: ONGOING, RATIONAL CALIBRATION OF REWARD-DRIVEN PERCEPTUAL BIASES

Yunshu Fan<sup>1</sup>, Joshua I. Gold<sup>1,2</sup> and Long Ding<sup>1,2</sup>

<sup>1</sup>Neuroscience Graduate Group, <sup>2</sup>Department of Neuroscience,  
Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Chapter 2 was published in *eLife*, published by Oxford University Press, as:  
Y. Fan, G. I. Gold, L. Ding. “Ongoing, rational calibration of reward-driven perceptual  
biases.” *eLife* 2018;7:e36018 DOI: 10.7554/eLife.36018

## Introduction

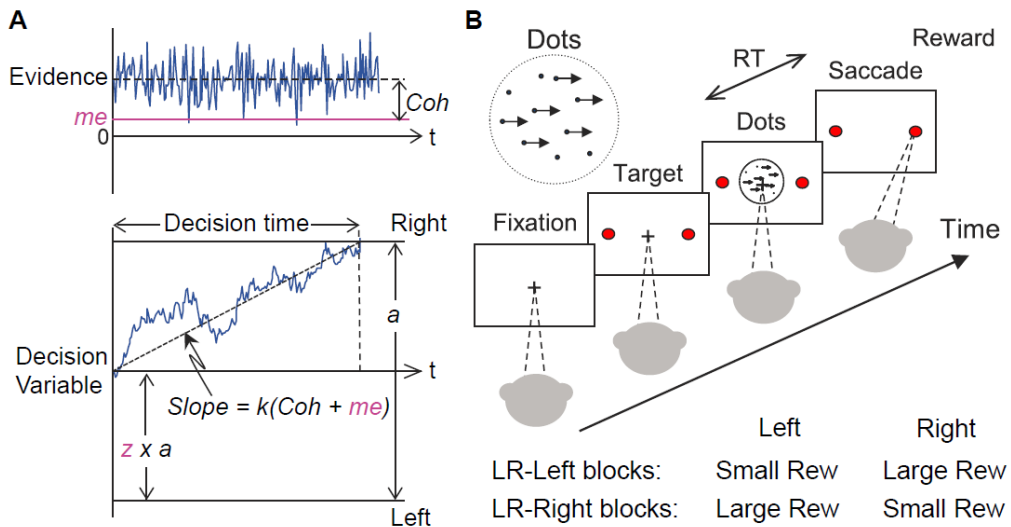
Normative theory has played an important role in our understanding of how the brain forms decisions. For example, many perceptual, memory, and reward-based decisions show inherent trade-offs between speed and accuracy. These trade-offs are parsimoniously captured by a class of sequential-sampling models, such as the drift-diffusion model (DDM), that are based on the accumulation of noisy evidence over time to a pre-defined threshold value, or bound (Ratcliff, 1978; Gold and Shadlen, 2002; Bogacz et al., 2006; Krajbich et al., 2010). These models have close ties to statistical decision theory, particularly the sequential probability ratio test that can, under certain assumptions, maximize expected accuracy for a given number of samples or minimize the number of samples needed for a given level of accuracy (Barnard, 1946; Wald, 1947; Wald and Wolfowitz, 1948). However, even when these models provide good descriptions of the average behavior of groups of subjects, they may not capture the substantial variability under different conditions and/or across individual subjects. The goal of this study was to better understand the principles that govern this variability, in particular how these principles relate to normative theory.

We focused on a perceptual decision-making task with asymmetric rewards. For this task, both human and animal subjects tend to make decisions that are biased towards the percept associated with the larger payoff (e.g. Maddox and Bohil, 1998; Voss et al., 2004; Diederich and Busemeyer, 2006; Liston and Stone, 2008; Serences, 2008; Feng et al., 2009; Simen et al., 2009; Nomoto et al., 2010; Summerfield and Koechlin, 2010; Teichert and Ferrera, 2010; Gao et al., 2011; Leite and Ratcliff, 2011; Mulder et al., 2012; Wang et al., 2013; White and Poldrack, 2014). These biases are roughly consistent with a rational strategy to maximize a particular reward function that

depends on both the speed and accuracy of the decision process, such as the reward rate per trial or per unit time (Gold and Shadlen, 2002; Bogacz et al, 2006). This strategy can be accomplished via context-dependent adjustments in a DDM-like decision process along two primary dimensions (Figure 2.1A): 1) the momentary sensory evidence, via the drift rate; and 2) the decision rule, via the relative bound heights that govern how much evidence is needed for each alternative (Ratcliff, 1985). Subjects tend to make adjustments along one or both of these dimensions to produce overall biases that are consistent with normative theory, but with substantial individual variability (Voss et al., 2004; Cicmil et al., 2015; Bogacz et al., 2006; Simen et al., 2009; Summerfield and Koechlin, 2010; Leite and Ratcliff, 2011; Mulder et al., 2012; Goldfarb et al., 2014).

To better understand the principles that govern these kinds of idiosyncratic behavioral patterns, we trained three monkeys to perform a response-time (RT), asymmetric-reward decision task with mixed perceptual uncertainty (Figure 2.1B). Like human subjects, the monkeys showed robust decision biases toward the large-reward option. These biases were sensitive to not just the reward asymmetry, as has been shown previously, but also to experience-dependent changes in perceptual sensitivity. These biases were consistent with adjustments to both the momentary evidence and decision rule in the DDM. However, these two adjustments favored the large- and small-reward choice, respectively, leading to nearly, but not exactly, maximal reward rates. We accounted for these adjustments in terms of a satisficing, gradient-based learning model that calibrated biases to balance the relative influence of perceptual and reward-based information on the decision process. Together, the results imply complementary roles of normative and heuristic principles to understand how the brain combines uncertain

sensory input and internal preferences to form decisions that can vary considerably across individuals and task conditions.



**Figure 2.1. Theoretical framework and task design.**

**(A)** Schematics of the drift-diffusion model (DDM). Motion evidence is modeled as samples from a unit-variance Gaussian distribution (mean: signed coherence,  $Coh$ ). Effective evidence is modeled as the sum of motion evidence and an internal momentary-evidence bias ( $me$ ). The decision variable starts at value  $a \times z$ , where  $z$  governs decision-rule bias, and accumulates effective evidence over time with a proportional scaling factor ( $k$ ). A decision is made when the decision variable reaches either bound. Reaction time (RT) is assumed to be the sum of the decision time and a saccade-specific non-decision time.

**(B)** Reaction-time (RT) random-dot visual motion direction discrimination task with asymmetric rewards. A monkey makes a saccade decision based on the perceived global motion of a random-dot kinematogram. Reward is delivered on correct trials and with a magnitude that depends on reward context. Two reward contexts (LR-Left and LR-Right) were alternated in blocks of trials with signaled block changes. Motion directions and strengths were randomly interleaved within blocks.

## Results

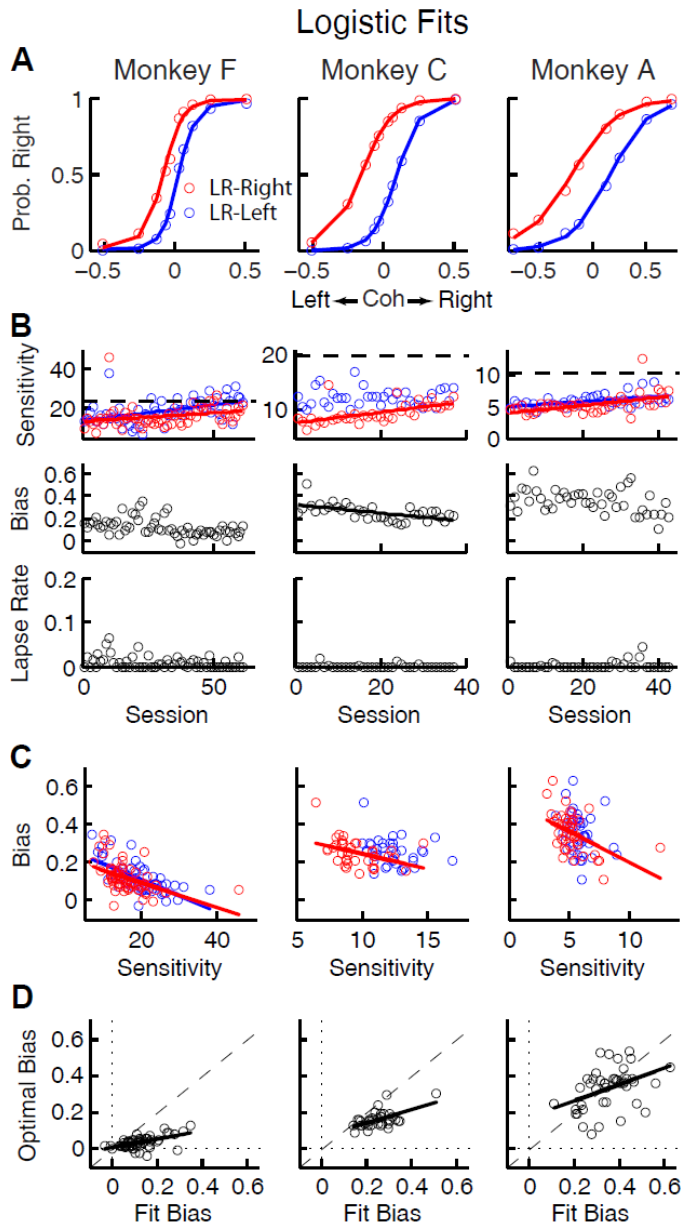
We trained three monkeys to perform the asymmetric-reward random-dot motion discrimination (“dots”) task (Figure 2.2A). All three monkeys were initially trained on a symmetric-reward version of the task for which they were required to make fast eye movements (saccades) in the direction congruent with the global motion of a random-dot kinematogram to receive juice reward. They then performed the asymmetric-reward versions that were the focus of this study. Specifically, in blocks of 30–50 trials, we alternated direction-reward associations between a “LR-Right” reward context (the large reward was paired with a correct rightward saccade and the small reward was paired with a correct leftward saccade) and the opposite “LR-Left” reward context. We also varied the ratio of large versus small reward magnitudes (“reward ratio”) across sessions for each monkey. Within a block, we randomly interleaved motion stimuli with different directions and motion strengths (expressed as coherence, the fraction of dots moving in the same direction). We monitored the monkey’s choice (which saccade to make) and RT (when to make the saccade) on each trial.

### ***The monkeys’ biases reflected changes in reward context and perceptual sensitivity***

For the asymmetric-reward task, all three monkeys tended to make more choices towards the large-reward option, particularly when the sensory evidence was weak. These choice biases corresponded to horizontal shifts in the psychometric function describing the probability of making a rightward choice as a function of signed motion coherence (negative for leftward motion, positive for rightward motion; Figure 2.2A, plus example fits shown in Figure 2.2–figure supplement 1). These functions showed

somewhat similar patterns of behavior but some differences in detail for the three monkeys. For example, each monkey showed steady increases in perceptual sensitivity (steepness of the psychometric function), which initially dropped relative to values from the symmetric-reward task then tended to increase with more experience with asymmetric rewards (Figure 2.2B, top;  $H_0$ : partial Spearman's  $\rho$  of sensitivity versus session index after accounting for session-specific reward ratios=0,  $p < 0.01$  in all cases, except LR-Left for monkey C, for which 0.56). Moreover, lapse rates were near zero across sessions (Figure 2.2B, bottom), implying that the monkeys knew how to perform the task. However, the monkeys differed in terms of overall bias, which was the smallest in monkey F. Nevertheless, for all three monkeys bias magnitude tended to decrease over sessions, although this tendency was statistically significant only for monkey C after accounting for co-variations with reward rate (Figure 2.2B, middle). There was often a negative correlation between choice bias and sensitivity, consistent with a general strategy of adjusting bias to obtain more reward (Figure 2.2C; Figure 2-figure supplement 2C). Monkeys F and C used suboptimal biases that were larger than the optimal values, whereas monkey A showed greater variations (Figure 2.2D). The monkeys showed only negligible or inconsistent sequential choice biases (Figure 2.2-figure supplement 1), and adding sequential terms did not substantially affect the best-fitting values of the non-sequential terms in the logistic regression (spearman's  $\rho > 0.8$  comparing session-by-session best-fitting values of the terms in Eq. (1) with and without additional sequential terms from Eq. (2)). Therefore, all subsequent analyses did not include sequential choice effects.





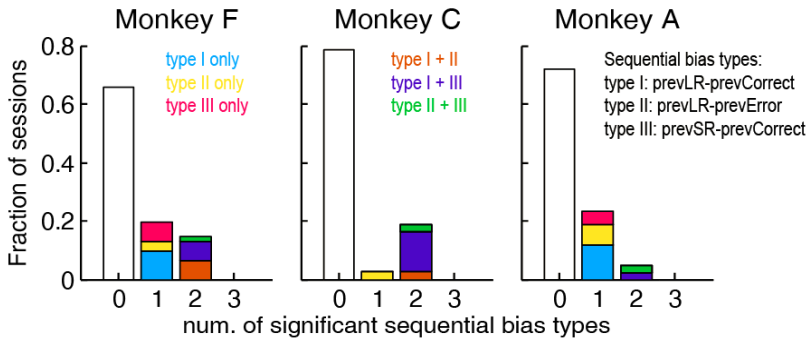
**Figure 2.2. Relationships between sensitivity and bias from logistic fits to choice data.**

(A) For each monkey, the probability of making a rightward choice is plotted as a function of signed coherence ( $-/+$  indicate left/right motion) from all sessions, separately for the two reward contexts, as indicated. Lines are logistic fits.

(B) Top row: Motion sensitivity (steepness of the logistic function) in each context as a function of session index (colors as in A). Solid lines indicate significant positive partial Spearman correlation after accounting for changes in reward ratio across sessions ( $p < 0.05$ ). Black dashed lines indicate each monkey's motion sensitivity for the task with equal rewards before training on this asymmetric reward task. Middle row:  $\Delta$ Bias (horizontal shift between the two psychometric functions for the two reward contexts at chance level) as a function of session index. Solid line indicates significant negative partial Spearman correlation after accounting for changes in reward ratio across sessions ( $p < 0.05$ ). Bottom row: Lapse rate as a function of session index (median=0 for all three monkeys).

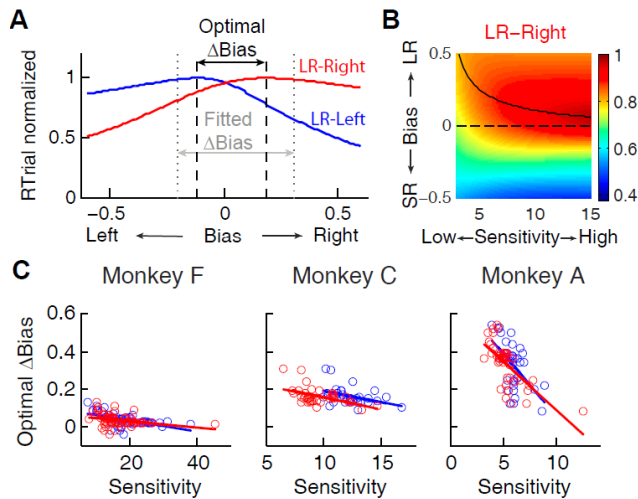
(C)  $\Delta$ Bias as a function of motion sensitivity for each reward context (colors as in A). Solid line indicates a significant negative partial Spearman correlation after accounting for changes in reward ratio across sessions ( $p < 0.05$ ).

(D) Optimal versus fitted  $\Delta$ bias. Optimal  $\Delta$ bias was computed as the difference in the horizontal shift in the psychometric functions in each reward context that would have resulted in the maximum reward per trial, given each monkey's fitted motion sensitivity and experienced values of reward ratio and coherences from each session (see Figure 2-figure supplement 2). Solid lines indicate significant positive Spearman correlations ( $p < 0.01$ ). Partial Spearman correlation after accounting for changes in reward ratio across sessions are also significant for monkeys F and C ( $p < 0.05$ ).



**Figure 2.2-figure supplement 1. Monkeys showed minimal sequential choice biases.**

Histogram of the fraction of sessions with 0, 1 or 2 types of sequential choice biases. Colors indicate the sequential bias types with respect to the previous reward (Large or Small) and outcome (Correct or Error), as indicated. Significant sequential bias effects were identified by a likelihood-ratio test for  $H_0$ ; the sequential term in the logistic regression=0,  $p < 0.05$ .



**Figure 2.2-figure supplement 2. The optimal bias decreases with increasing sensitivity.**

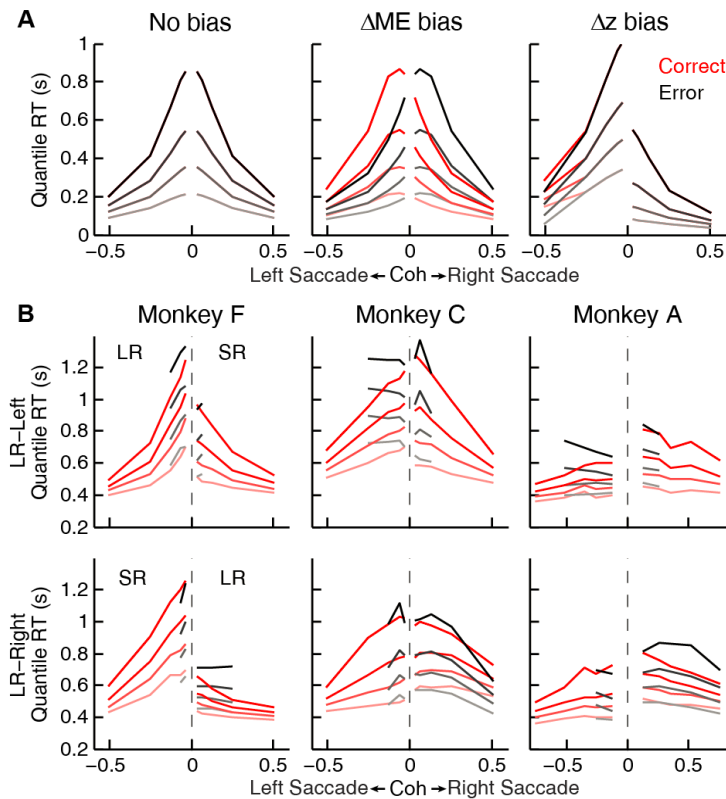
**(A)** Identification of the optimal  $\Delta$ bias for an example session using logistic fits. For each reward context (blue for LR-Left and red for LR-Right), RTrial was computed as a function of bias values sampled uniformly over a broad range, given the session-specific sensitivities, lapse rate, coherences and large:small reward ratio. The optimal  $\Delta$ bias was defined as the difference between the bias values with the maximal RTrial for the two reward contexts. The fitted  $\Delta$ bias was defined as the difference between the fitted bias values for the two reward contexts.

**(B)** The optimal bias decreases with increasing sensitivity. The example heatmap shows normalized RTrial as a function of sensitivity and bias values in the LR-Right blocks, assuming the same coherence levels as used for the monkeys and a large:small reward ratio of 2.3. The black curve indicates the optimal bias values for a given sensitivity value.

**(C)** Scatterplots of optimal  $\Delta$ biases obtained via the procedure described above as a function of sensitivity for each of the two reward contexts. Same format as Figure 3B. Solid lines indicate significant partial Spearman correlation after accounting for changes in reward ratio across sessions ( $p < 0.05$ ). Note that the scatterplots of the monkeys'  $\Delta$ biases and sensitivities in Figure 2C also show negative correlations, similar to this pattern.

To better understand the computational principles that governed these idiosyncratic biases, while also taking into account systematic relationships between the choice and RT data, we fit single-trial RT data (i.e., we modeled full RT distributions, not just mean RTs) from individual sessions to a DDM. We used a hierarchical-DDM (HDDM) method that assumes that parameters from individual sessions of the same

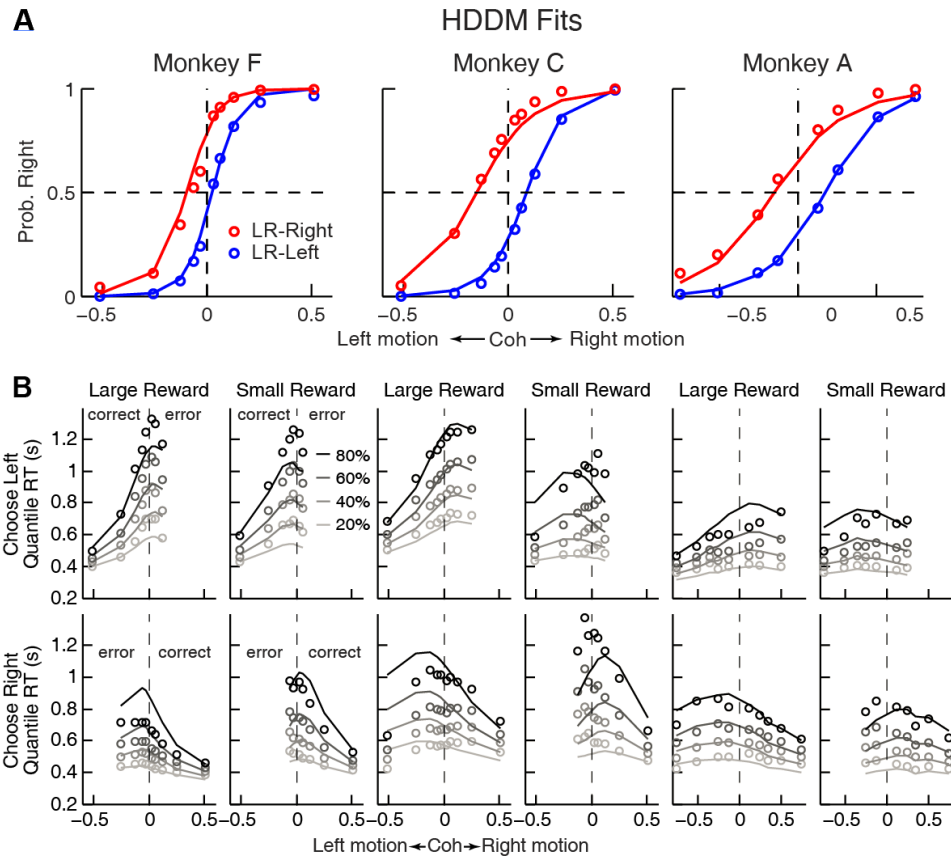
monkey are samples from a group distribution (Wiecki et al., 2013). The HDDM was fit to data from each monkey separately. The HDDM had six parameters for each reward context. Four were from a basic DDM (Figure 2.1A):  $a$ , the total bound height, representing the distance between the two choice bounds;  $k$ , a scaling factor that converts sensory evidence (motion strength and direction) to the drift rate; and  $t_0$  and  $t_1$ , non-decision times for leftward and rightward choices, respectively. The additional two parameters provided biases that differed in terms of their effects on the full RT distributions (Figure 2.3–figure supplement 1):  $me$ , which is additional momentary evidence that is added to the motion evidence at each accumulating step and has asymmetric effects on the two choices and on correct versus error trials (positive values favor the rightward choice); and  $z$ , which determines the decision rules for the two choices and tends to have asymmetric effects on the two choices but not on correct versus error trials (values  $>0.5$  favor the rightward choice). The HDDM fitting results are shown in Figure 2.3, and summaries of best-fitting parameters and goodness-of-fit metrics are provided in Table 1. A DDM variant with collapsing bounds provided qualitatively similar results as the HDDM (Figure 2.3–figure supplement 2). Thus, subsequent analyses use the model with fixed bounds, unless otherwise noted.



**Figure 2.3–figure supplement 1. Qualitative comparison between the monkeys' RT distribution and DDM predictions.**

**(A)** RT distributions as predicted by a DDM with no bias in decision rule ( $z$ ) or momentary evidence ( $me$ ; left), with  $me > 0$  (middle), and with  $z > 0.5$  (right). RT distributions are shown separately for correct (red) and error (black) trials and using values corresponding to 20th, 40th, 60th, and 80th percentiles. Note that the predictions assumed zero non-decision time to demonstrate effects on RT by only  $me$  or  $z$  biases. Positive/negative coh values indicate rightward/leftward saccades. The values of  $me$  and  $z$  were chosen to induce similar choice biases ( $\sim 0.075$  in coherence units). Note that the  $me$  bias induces large asymmetries in RT both between the two choices and between correct and error trials, whereas the  $z$  bias induces a large asymmetry in RT for the two choices, but with little asymmetry between correct and error trials.

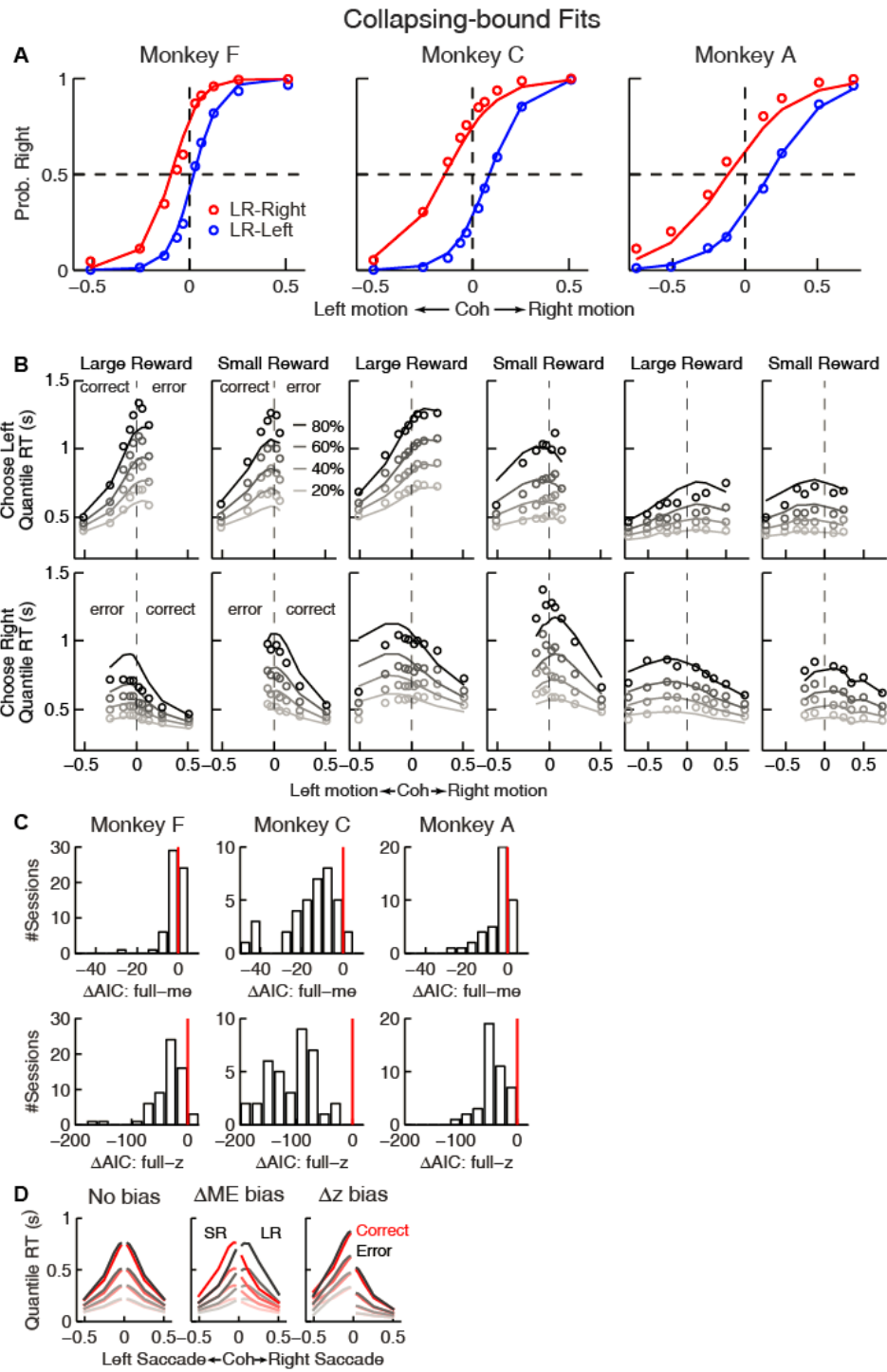
**(B)** The monkeys' mean RTs for four quantiles for the LR-Right (*top*) and LR-Left (*bottom*) reward contexts, respectively (same convention as in A). Note the presence of substantial asymmetries between correct and error trials for all three monkeys.



**Figure 2.3. Comparison of choice and RT data to HDDM fits with both momentary-evidence (*me*) and decision-rule (*z*) biases.**

**(A)** Psychometric data (points as in Figure 2A) shown with predictions based on HDDM fits to both choice and RT data.

**(B)** RT data (circles) and HDDM-predicted RT distributions (lines). Both sets of RT data were plotted as the session-averaged values corresponding to the 20<sup>th</sup>, 40<sup>th</sup>, 60<sup>th</sup>, and 80<sup>th</sup> percentiles of the full distribution for the five most frequently used coherence levels (we only show data when >40% of the total sessions contain >4 trials for that combination of motion direction, coherence and reward context). Top row: Trials in which monkey chose the left target. Bottom row: Trials in which monkeys chose the right target. Columns correspond to each monkey (as in A), divided into choices in the large- (left column) or small- (right column) reward direction (correct/error choices are as indicated in the left-most columns; note that no reward was given on error trials). The HDDM-predicted RT distributions were generated with 50 runs of simulations, each run using the number of trials per condition (motion direction × coherence × reward context × session) matched to experimental data and using the best-fitting HDDM parameters for that monkey



**Figure 2.3—figure supplement 2. Fits to a DDM with collapsing bounds.**

**(A, B)** A DDM with collapsing bounds and both momentary evidence ( $me$ ) and decision rule ( $z$ ) biases fit to each monkey's RT data. Same format as Figure 3.

**(C)** The model that included both  $me$  and  $z$  adjustments ("full") had smaller Akaike Information Criterion (AIC) values than reduced models (" $me$ " or " $z$ " only) across sessions. Note also the different ranges of  $\Delta AIC$  for the full- $me$  and full- $z$  comparisons. The mean  $\Delta AIC$  (full- $me$ ) and  $\Delta AIC$  (full- $z$ ) values are significantly different from zero (Wilcoxon signed rank test,  $p=0.0007$  for Monkey F's full- $me$  comparison and  $p<0.0001$  for all others).

**(D)** RT distributions as predicted by the DDM with collapsing bounds, using no bias in  $z$  or  $me$  (left),  $me>0$  (middle), or  $z>0.5$  (right). Same format as Figure 3—figure supplement 1A.

**Table 2.1 Best fitting DDM parameters.**

	Monkey F (26079 trials)				Monkey C (37161 trials)				Monkey F (21089 trials)			
	LR-Left		LR-Right		LR-Left		LR-Right		LR-Left		LR-Right	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
$a$	1.67	0.16	1.43	0.12	1.77	0.09	1.53	0.13	1.33	0.13	1.36	0.09
$k$	10.22	1.87	9.91	2.11	6.58	0.51	5.08	0.92	4.04	0.33	3.45	0.46
$t_1$	0.31	0.03	0.29	0.03	0.35	0.04	0.33	0.05	0.29	0.04	0.27	0.04
$t_0$	0.28	0.04	0.31	0.05	0.33	0.04	0.31	0.03	0.21	0.08	0.26	0.04
$z$	0.60	0.03	0.57	0.04	0.62	0.03	0.40	0.04	0.57	0.06	0.39	0.04
$me$	-0.06	0.04	0.08	0.05	-0.14	0.04	0.21	0.06	-0.22	0.05	0.27	0.09

The DDM fits provided a parsimonious account of both the choice and RT data. Consistent with the results from the logistic analyses, the HDDM analyses showed that the monkeys made systematic improvements in psychometric sensitivity ( $H_0$ : partial Spearman's  $\rho$  of sensitivity versus session index after accounting for session-specific reward ratios=0,  $p<0.01$  in all cases except  $p=0.06$  for LR-Left for monkey A). Moreover, there was a negative correlation between psychometric sensitivity and choice bias ( $H_0$ : partial Spearman's  $\rho$  of sensitivity versus total bias after accounting for session-specific



reward ratios=0,  $p < 0.001$  in all cases). These fits ascribed the choice biases to changes in both the momentary evidence ( $me$ ) and the decision rule ( $z$ ) of the decision process, as opposed to either parameter alone (Table 2). These fits also indicated context-dependent differences in non-decision times, which were smaller for all large-reward choices for all three monkeys except in the LR-Right context for monkeys C and A ( $t$ -test,  $p < 0.05$ ). However, the differences in non-decision times were relatively small across reward contexts, suggesting that the observed reward biases were driven primarily by effects on decision-related processes.

**Table 2.2 Model comparisons.**

The difference in deviance information criterion ( $\Delta$ DIC) between the full model (i.e., the model that includes both  $me$  and  $z$ ) and either reduced model ( $me$ -only or  $z$ -only), for experimental data and data simulated using each reduced model. Negative/positive values favor the full/reduced model. Note that the  $\Delta$ DIC values for the experimental data were all strongly negative, favoring the full model. In contrast, the  $\Delta$ DIC values for the simulated data were all positive, implying that this procedure did not simply prefer the more complex model.

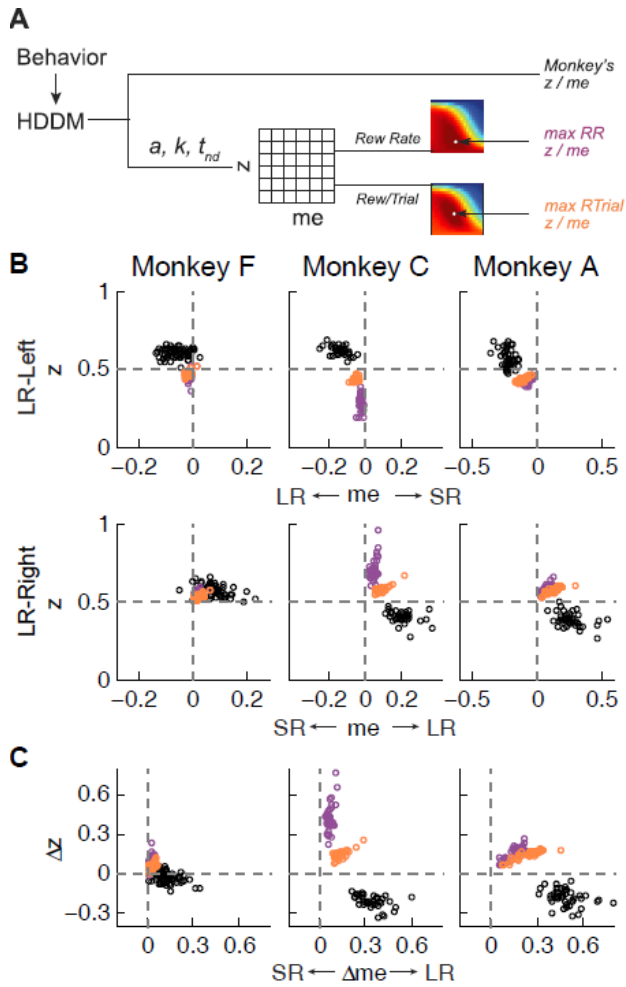
	Experimental Data				Simu: $me$ model		Simu: $z$ model	
	$\Delta$ DIC: full – $me$		$\Delta$ DIC: full – $z$		$\Delta$ DIC: full – $me$		$\Delta$ DIC: full – $z$	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Monkey F	-124.6	2.3	-2560.4	5.2	3.1	9.8	0.2	11.8
Monkey C	-1700.4	2.1	-6937.9	1.3	17.5	11.3	1.8	1.3
Monkey A	-793.6	3.4	-2225.7	4.0	25.4	9.0	1.2	3.4

***The monkeys' bias adjustments were adaptive with respect to optimal reward-rate functions***

To try to identify common principles that governed these monkey- and context-dependent decision biases, we analyzed behavior with respect to optimal benchmarks based on certain reward-rate functions. We focused on reward per unit time (RR) and

per trial (RTrial), which for this task are optimized in a DDM framework by adjusting momentary-evidence ( $me$ ) and decision-rule ( $z$ ) biases, such that both favor the large-reward choice. However, the magnitudes of these optimal adjustments depend on other task parameters ( $a$ ,  $k$ ,  $t_0$ , and  $t_1$ , non-bias parameters from the DDM, plus the ratio of the two reward sizes and inter-trial intervals) that can vary from session to session. Thus, to determine the optimal adjustments, we performed DDM simulations with the fitted HDDM parameters from each session, using different combinations of  $me$  and  $z$  values (Figure 2.4A). As reported previously (Bogacz et al., 2006; Simen et al., 2009), when the large reward was paired with the leftward choice, the optimal strategy used  $z < 0.5$  and  $me < 0$  (Figure 2.4B, top panels, purple and orange circles for RR and RTrial, respectively). Conversely, when the larger reward was paired with the rightward choice, the optimal strategy used  $z > 0.5$  and  $me > 0$  (Figure 2.4B, bottom panels).

The monkeys' adjustments of momentary-evidence ( $me$ ) and decision-rule ( $z$ ) biases showed both differences and similarities with respect to these optimal predictions (Figure 2.4B, black circles; similar results were obtained using fits from a model with collapsing bounds, Figure 2.4–figure supplement 1). In the next section, we consider the differences, in particular the apparent use of shifts in  $me$  in the adaptive direction (i.e., favoring the large-reward choice) but of a magnitude that was larger than predicted, along with shifts in  $z$  that tended to be in the non-adaptive direction (i.e., favoring the small-reward choice). Here we focus on the similarities and show that the monkeys' decision biases were adaptive with respect to the reward-rate function in four ways (RTrial provided slightly better predictions of the data and thus are presented in the main figures; results based on RR are presented in the Supplementary Figures).

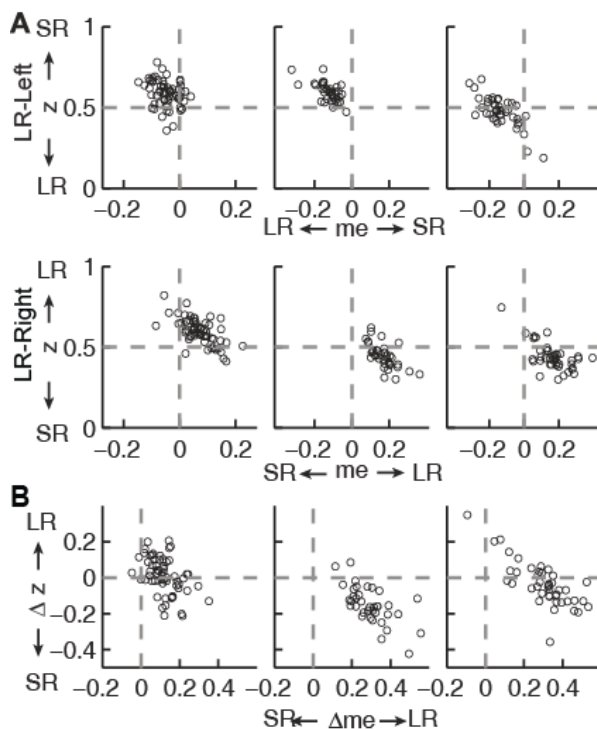


**Figure 2.4. Actual versus optimal adjustments of momentary-evidence ( $me$ ) and decision-rule ( $z$ ) biases.**

(A) Schematic of the comparison procedure. Choice and RT data from the two reward contexts in a given session were fitted separately using the HDDM. These context- and session-specific best-fitting  $me$  and  $z$  values are plotted as the monkey's data (black circles in B and C). Optimal values were determined by fixing parameters  $a$ ,  $k$ , and non-decision times at best-fitted values from the HDDM and searching in the  $me/z$  grid space for combinations of  $me$  and  $z$  that produced maximal reward function values. For each  $me$  and  $z$  combination, the predicted probability of left/right choice and RTs were used with actual task information (inter-trial interval, error timeout and reward sizes) to calculate the expected reward rate (RR) and average reward per trial (RTrial). Optimal  $me/z$  adjustments were then identified to maximize RR (purple) or RTrial (orange).

**(B)** Scatterplots of the monkeys'  $me/z$  adjustments (black), predicted optimal adjustments for maximal RR (purple), and predicted optimal adjustments for maximal Rtrial (orange), for the two reward contexts in all sessions (each data point was from a single session). Values of  $me > 0$  or  $z > 0.5$  produce biases favoring rightward choices.

**(C)** Scatterplots of the differences in  $me$  (abscissa) and  $z$  (ordinate) between the two reward contexts for monkeys (black), for maximizing RR (purple), and for maximizing Rtrial (orange). Positive  $\Delta me$  and  $\Delta z$  values produce biases favoring large-reward choices.



**Figure 2.4—figure supplement 1. Estimates of momentary-evidence ( $me$ ) and decision-rule ( $z$ ) biases using the collapsing-bound DDM fits.**

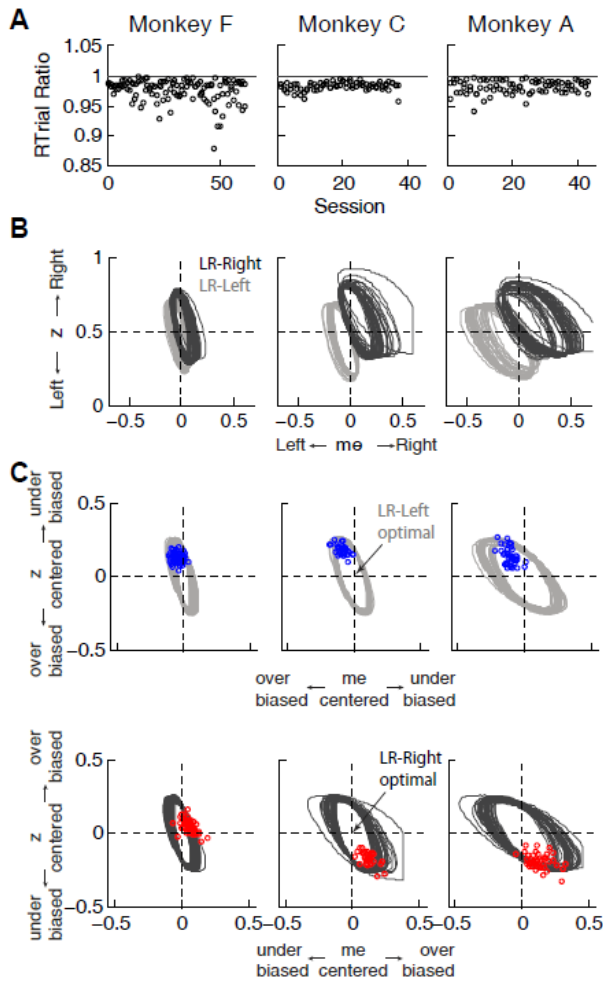
Same format as Figure 4B and C, except here only showing fits to the monkeys' data. As with the model without collapsing bounds, the adjustments in  $me$  tended to favor the large reward but the adjustments in  $z$  tended to favor the small reward.

First, the best-fitting  $me$  and  $z$  values from each monkey corresponded to near-maximal reward rates (Figure 2.5A). We compared the optimal values of reward per trial ( $R_{\text{Trial}_{\text{max}}}$ ) to the values predicted from the monkeys' best-fitting  $me$  and  $z$  adjustments ( $R_{\text{Trial}_{\text{predict}}}$ ). Both  $R_{\text{Trial}_{\text{predict}}}$  and  $R_{\text{Trial}_{\text{max}}}$  depended on the same non-bias parameters in the HDDM fits that were determined per session ( $a$ ,  $k$ ,  $t_0$ , and  $t_1$ ) and thus are directly

comparable. Their ratios tended to be nearly, but slightly less than, one (mean ratio: 0.977, 0.984, and 0.983 for monkeys F, C, and A, respectively) and remained relatively constant across sessions ( $H_0$ : slopes of linear regressions of these ratios versus session number=0,  $p>0.05$  for all three monkeys). Similar results were also obtained using the monkeys' realized rewards, which closely matched  $R_{\text{Trial}_{\text{predict}}}$  (mean ratio: 0.963, 0.980 and 0.974; across-session Spearman's  $\rho=0.976$ , 0.995, and 0.961, for monkeys F, C, and A, respectively,  $p<0.0001$  in all three cases). These results reflected the shallow plateau in the  $R_{\text{Trial}}$  function near its peak (Figure 2.5B), such that the monkeys' actual adjustments of  $me$  and  $z$  were within the contours for 97%  $R_{\text{Trial}_{\text{max}}}$  in most sessions (Figure 2.5C; see Figure 2.5–figure supplement 1 for results using RR). Thus, the monkeys' overall choice biases were consistent with strategies that lead to nearly optimal reward outcomes.

Second, the across-session variability of each monkey's decision biases was predicted by idiosyncratic features of the reward functions. The reward functions were, on average, different for the two reward contexts and each of the three monkeys (Figure 6A). These differences included the size of the near-maximal plateau (red patch), which determined the level of tolerance in  $R_{\text{Trial}}$  for deviations from optimal adjustments in  $me$  and  $z$ . This tolerance corresponded to the session-by-session variability in each monkey's  $me$  and  $z$  adjustments (Figure 2.6B). In general, monkey F had the smallest plateaus and tended to use the narrowest range of  $me$  and  $z$  adjustments across sessions. In contrast, monkey A had the largest plateaus and tended to use the widest range of  $me$  and  $z$  adjustments (Pearson's  $\rho$  between the size of the 97%  $R_{\text{Trial}}$  contour, in pixels, and the sum of the across-session variances in each monkeys'  $me$

and  $z$  adjustments=0.83,  $p=0.041$ ). Analyses using the RR function produced qualitatively similar results (Figure 2.6–figure supplement 1).

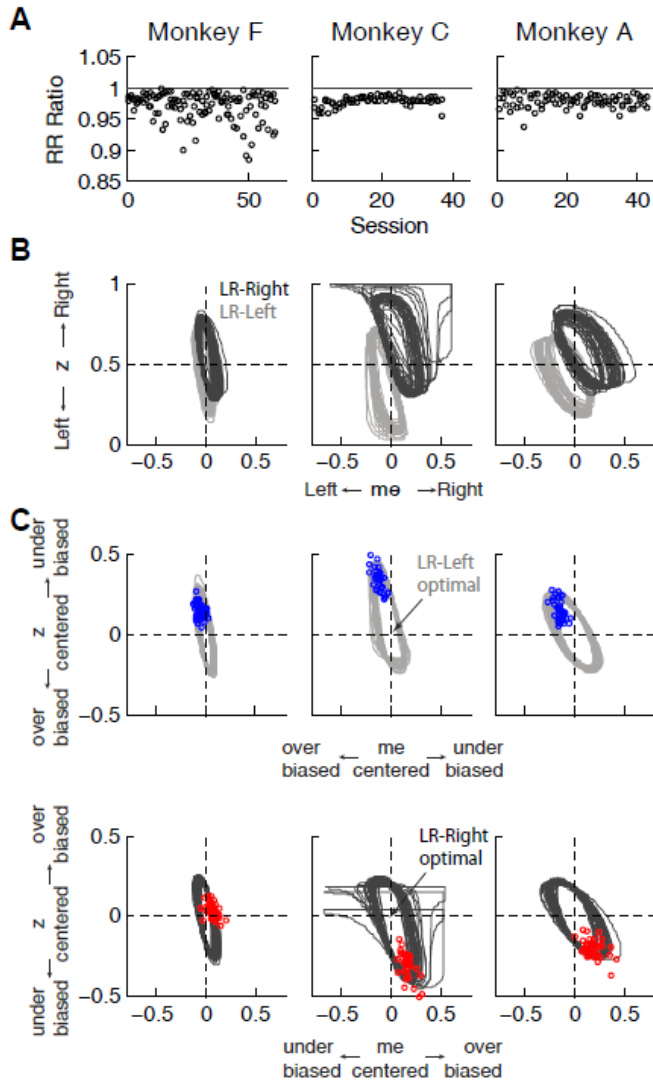


**Figure 2.5. Predicted versus optimal reward per trial (RTrial).**

**(A)** Scatterplots of  $R_{\text{Trial}}^{\text{predict}}:R_{\text{Trial}}^{\text{max}}$  ratio as a function of session index. Each session was represented by two ratios, one for each reward context. Mean ratio across contexts and sessions: 0.977 for monkey F, 0.984 for monkey C, and 0.983 for monkey A.

**(B)** 97%  $R_{\text{Trial}}^{\text{max}}$  contours for all sessions, computed using the best-fitting HDDM parameters and experienced coherences and reward ratios from each session. Light grey: LR-Left blocks; Dark grey: LR-Right blocks.

**(C)** The monkeys' adjustments (blue in LR-Left blocks, red in LR-Right blocks) were largely within the 97%  $R_{\text{Trial}}^{\text{max}}$  contours for all sessions and tended to cluster in the *me* over-biased, *z* under-biased quadrants (except Monkey F in the LR-Right blocks). The contours and monkeys' adjustments are centered at the optimal adjustments for each session.



**Figure 2.5–figure supplement 1.**  
**Predicted versus optimal reward rate (RR).** Same format as Figure 5.  
 Mean  $RR_{\text{predict}}:RR_{\text{max}}$  ratio across sessions=0.971 for monkey F, 0.980 for monkey C, and 0.980 for monkey A.





**Figure 2.6. Relationships between adjustments of momentary-evidence ( $me$ ) and decision-rule ( $z$ ) biases and RTrial function properties.**

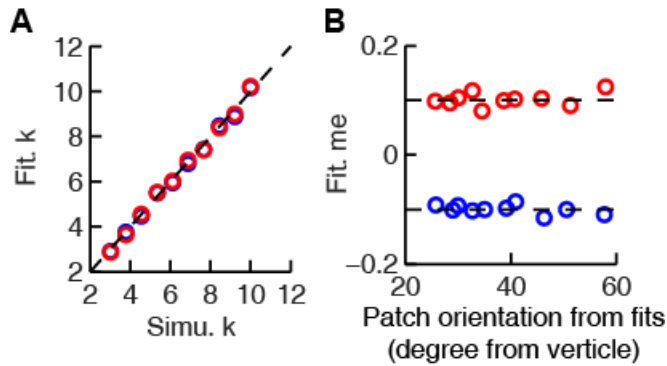
**(A)** Mean RTrial as a function of  $me$  and  $z$  adjustments for the LR-Left (top) and LR-Right (bottom) blocks. Hotter colors represent larger RTrial values (see legend to the right). RTrial was normalized to RTrial<sub>max</sub> for each session and then averaged across sessions.

**(B)** Scatterplot of the total variance in  $me$  and  $z$  adjustments across sessions (ordinate) and the area of >97% max of the average RTrial patch (abscissa). Variance and patch areas were measured separately for the two reward blocks (circles for LR-Left blocks, squares for LR-Right blocks).

**(C, D)** The monkeys' session- and context-specific values of  $me$  (C) and  $z$  (D) co-varied with the orientation of the >97% heatmap patch (same as the contours in Figure 5B). Orientation is measured as the angle of the tilt from vertical. Circles: data from LR-Left block; squares: data from LR-Right block; lines: significant correlation between  $me$  (or  $z$ ) and patch orientations across monkeys ( $p < 0.05$ ). Colors indicate different monkeys (see legend in B).

**(E)** Scatterplots of conditionally optimal versus fitted  $\Delta me$  (top row) and  $\Delta z$  (bottom row). For each reward context, the conditionally optimal  $me$  ( $z$ ) value was identified given the monkey's best-fitting  $z$  ( $me$ ) values. The conditionally optimal  $\Delta me$  ( $\Delta z$ ) was the difference between the two conditional optimal  $me$  ( $z$ ) values for the two reward contexts. Grey lines indicate the range of conditional  $\Delta me$  ( $\Delta z$ ) values corresponding to the 97% maximal RTrial given the monkeys' fitted  $z$  ( $me$ ) values.





**Figure 2.6–figure supplement 2: The HDDM model fitting procedure does not introduce spurious correlations between patch orientation and *me* value.** Artificial sessions were simulated with fixed *me* values ( $\pm 0.1$  for the two reward contexts) and different *k* values.

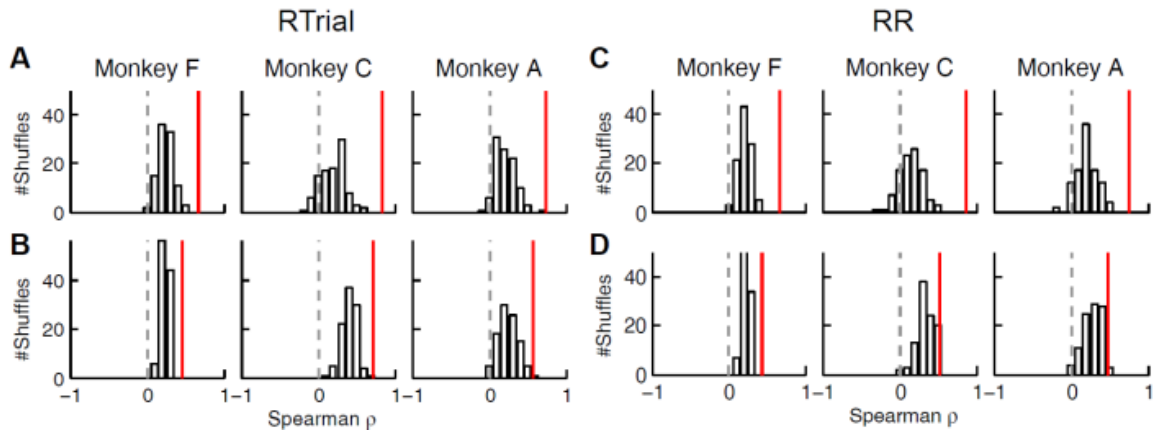
**(A)** Recovered *k* values from HDDM fitting closely matched *k* values used for the simulations.

**(B)** Recovered *me* values from HDDM fitting closely matched *me* values used for simulation and did not correlate with RTrial patch orientation.

Third, the session-by-session adjustments in both *me* and *z* corresponded to particular features of each monkey’s context-specific reward function. The shape of this function, including the orientation of the plateau with respect to *z* and *me*, depended on the monkey’s perceptual sensitivity and the reward ratio for the given session. The monkeys’ *me* and *z* adjustments varied systematically with this orientation (Figure 2.6C and D for RTrial, Figure 2.6–figure supplement 1C and D for RR). This result was not an artifact of the fitting procedure, which was able to recover appropriate, simulated bias parameter values regardless of the values of non-bias parameters that determine the shape of the reward function (Figure 2.6–figure supplement 2).

Fourth, the monkeys’ *me* and *z* adjustments were correlated with the values that would maximize RTrial, given the value of the other parameter for the given session and reward context (Figure 2.6E for RTrial, Figure 2.6–figure supplement 1E for RR). These correlations were substantially weakened by shuffling the session-by-session reward

functions (Figure 2.6–figure supplement 3). Together, these results suggest that all three monkeys used biases that were adaptively calibrated with respect to the reward information and perceptual sensitivity of each session.



**Figure 2.6–figure supplement 3. The correlation between fitted and conditionally optimal adjustments was stronger for the real, session-by-session data (red lines) than for unmatched (shuffled) sessions (bars).**

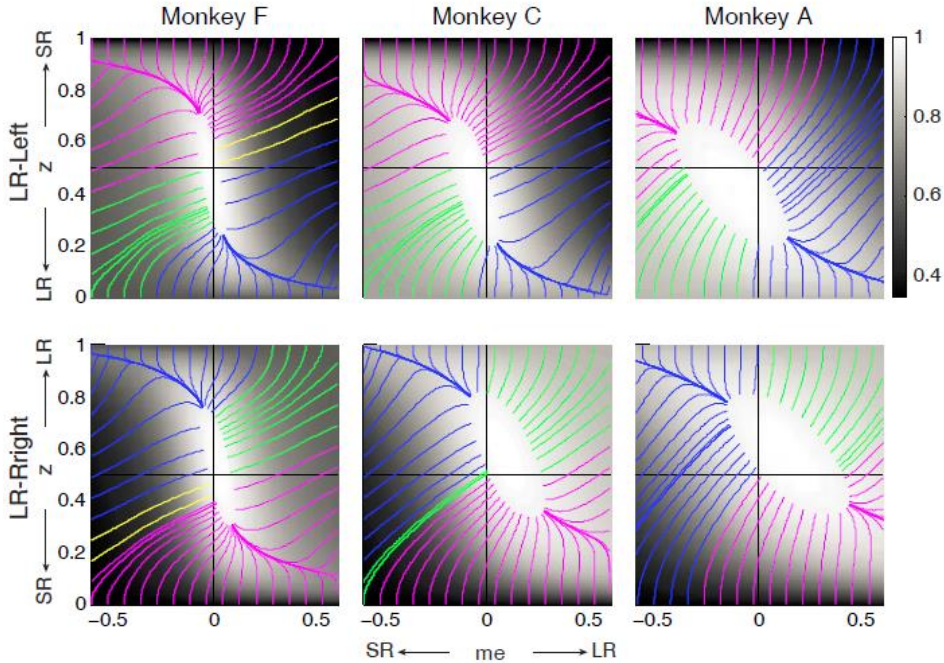
**(A, C)** Momentary-evidence ( $\Delta me$ ) adjustments. **(B, D)** Decision-rule ( $\Delta z$ ) adjustments. A, B: optimal values obtained with the R-Trial function. C, D: optimal values obtained with the RR function. Red lines indicate the partial Spearman correlation coefficients between the fitted and optimal  $\Delta me$  or  $\Delta z$  (obtained in the same way as data in Figure 6E) for matched sessions. Bars represent the histograms of partial correlation for unmatched sessions, which were obtained by 100 random shuffles of the sessions (i.e., comparing the optimal and best-fitting values from different sessions). Note that the histograms for the unmatched sessions are centered at positive values, reflecting the non-session-specific tendency of reward surfaces to skew towards overly biased  $me$  and  $z$  values. The correlation values for matched sessions (red lines) are at even more positive values (Wilcoxon rank-sum test,  $p < 0.001$  for all three monkeys and both  $\Delta me$  and  $\Delta z$ ), suggesting additional session-specific tuning of the  $me$  and  $z$  parameters.

***The monkeys’ adaptive adjustments were consistent with a satisficing, gradient-based learning process***

Thus far, we showed that all three monkeys adjusted their decision strategies in a manner that matched many features of the optimal predictions based on their

idiosyncratic, context-specific reward-rate functions. However, their biases did not match the optimal predictions exactly. Specifically, all three monkeys used shifts in  $me$  favoring the large-reward choice (adaptive direction) but of a magnitude that was larger than predicted, along with shifts in  $z$  favoring the small-reward choice (non-adaptive direction). We next show that these shifts can be explained by a model in which the monkeys are initially over-biased, then adjust their model parameters to increase reward and stop learning when the reward is high enough, but not at its maximum possible value.

The intuition for this gradient-based satisficing model is shown in Figure 2.7. The lines on the  $R_{\text{Trial}}$  heatmap represent the trajectories of a gradient-tracking procedure that adjusts  $me$  and  $z$  values to increase  $R_{\text{Trial}}$  until a termination point (for illustration, here we used 97% of the maximum possible value). Gradient lines are color-coded based on how  $me$  and  $z$  values at the end points relate to the optimal  $me$  and  $z$  values. For example, consider adjusting  $me$  and  $z$  by following all of the magenta gradient lines until their endpoints. The lines are color-coded by  $me/z$  being adaptive vs. non-adaptive, regardless of their relative magnitudes to the optimal values. In other words, as long as the initial  $me$  and  $z$  values fall within the area covered by the magenta lines, the positive gradient-tracking procedure would lead to a good-enough solution with over-shifted  $me$  and non-adaptive  $z$  values similar to what we found in the monkeys' data. Figure 2.7 also illustrates why assumptions about the starting point of this adaptive process are important: randomly selected starting points would result in learned  $me$  and  $z$  values distributed around the peak of the reward function, whereas the data (e.g., Figure 2.5C) show distinct clustering that implies particular patterns of starting points.



**Figure 2.7. Relationships between starting and ending values of the satisficing, reward function gradient-based updating process.**

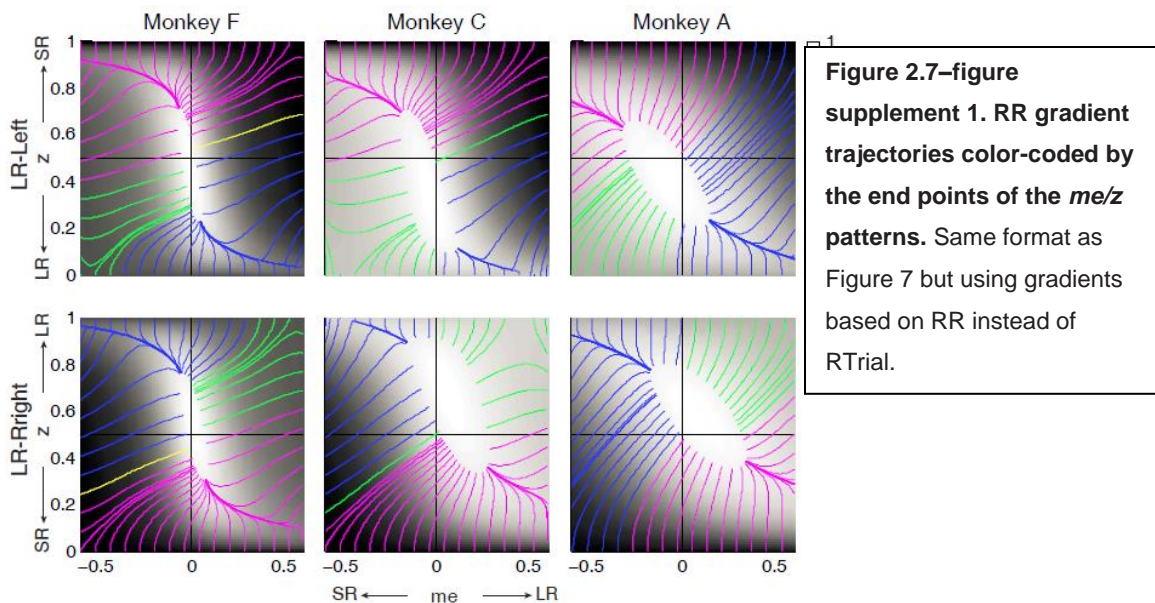
Example gradient lines of the average RTrial maps for the three monkeys are color coded based on the end point of gradient-based  $me$  and  $z$  adjustments in the following

ways: 1)  $me$  biases to large reward whereas  $z$  biases to small reward (magenta); 2)  $z$  biases to large reward whereas  $me$  biases to small reward (blue); 3)  $me$  and  $z$  both bias to large reward (green), and 4)  $me$  and  $z$  both bias to small reward (yellow). The gradient lines ended on the 97%  $RTrial_{max}$  contours.

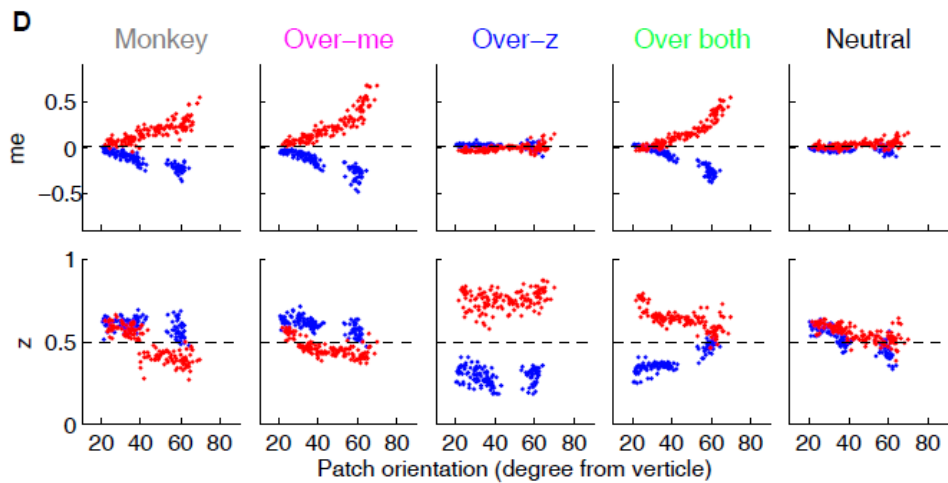
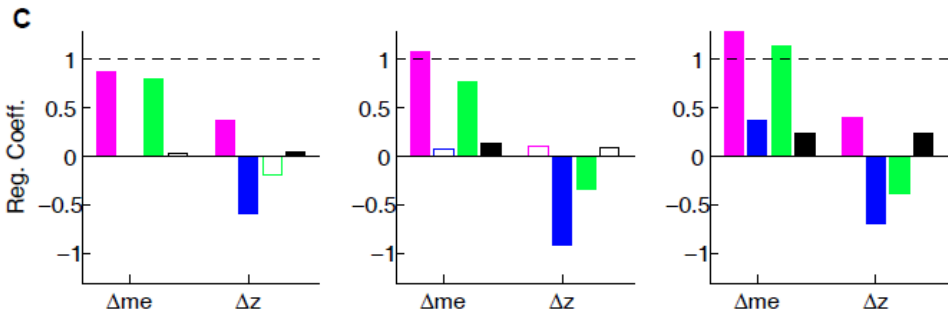
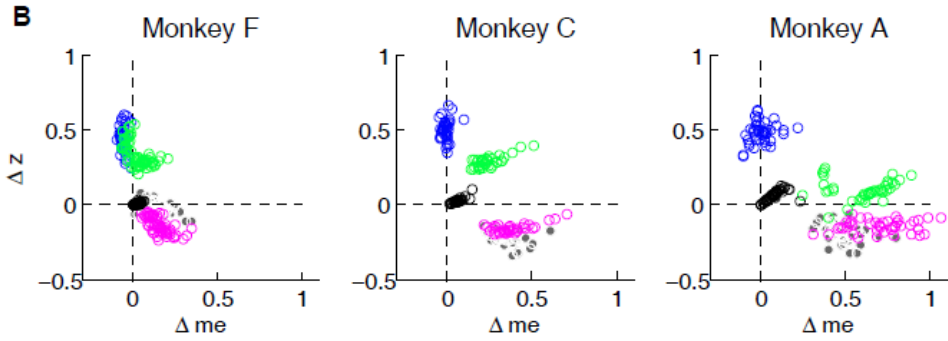
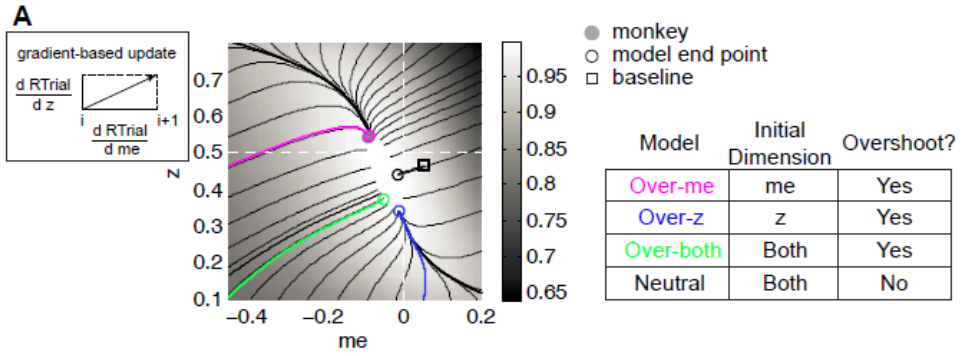
Top row: LR-Left block; bottom row: LR-Right block.

We simulated this process using: 1) different starting points; 2) gradients defined by the reward function derived separately for each reward context, session, and monkey; and 3) a termination rule corresponding to achieving each monkey's average reward in that session ( $RTrial_{predict}$ ) estimated from the corresponding best-fitting model parameters and task conditions. This process is illustrated for LR-Left blocks in an example session from monkey C (Figure 2.8A). We estimated the unbiased  $me$  and  $z$  values as the midpoints between their values for LR-Left and LR-Right blocks (square). At this point,

the RTrial gradient is larger along the *me* dimension than the *z* dimension, reflecting the tilt of the reward function. We set the initial point at baseline *z* and a very negative value of *me* (90% of the highest coherence used in the session; overshoot in the adaptive direction) and referred to this setting as the “over-*me*” model. The *me* and *z* values were then updated according to the RTrial gradient (see cartoon insert in Figure 2.8A), until the monkey’s  $RTrial_{predict}$  or better was achieved (magenta trace and circle). The endpoint of this updating process was very close to monkey C’s actual adjustment (gray circle). For comparison, three alternative models are illustrated. The “over-*z*” model selects *z* as the initial dimension and assumes updating from the baseline *me* and over-adjusted *z* values (blue, initial *z* set as 0.1 for the LR-Left context and 0.9 for the LR-Right context). The “over-both” model assumes updating from the over-adjusted *me* and *z* values (green). The “neutral” model assumes the same updating process but from the baseline *me* and baseline *z* (black). The endpoints from these alternative models deviated considerably from the monkey’s actual adjustment.









**Figure 2.8. The satisficing reward function gradient-based model.**

A, Illustration of the procedure for predicting a monkey's  $me$  and  $z$  values for a given RTrial function. For better visibility, RTrial for the LR-Left reward context in an example session is shown as a heatmap in greyscale. Gradient lines are shown as black lines. The square indicates the unbiased  $me$  and  $z$  combination (average values across the two reward contexts). The four trajectories represent gradient-based searches based on four alternative assumptions of initial values (see table on the right). All four searches stopped when the reward exceeded the average reward the monkey received in that session ( $RTrial_{predict}$ ), estimated from the corresponding best-fitting model parameters and task conditions. Open circles indicate the end values. Grey filled circle indicates the monkey's actual  $me$  and  $z$ . Note that the end points differ among the four assumptions, with the magenta circle being the closest to the monkey's fitted  $me$  and  $z$  of that session.

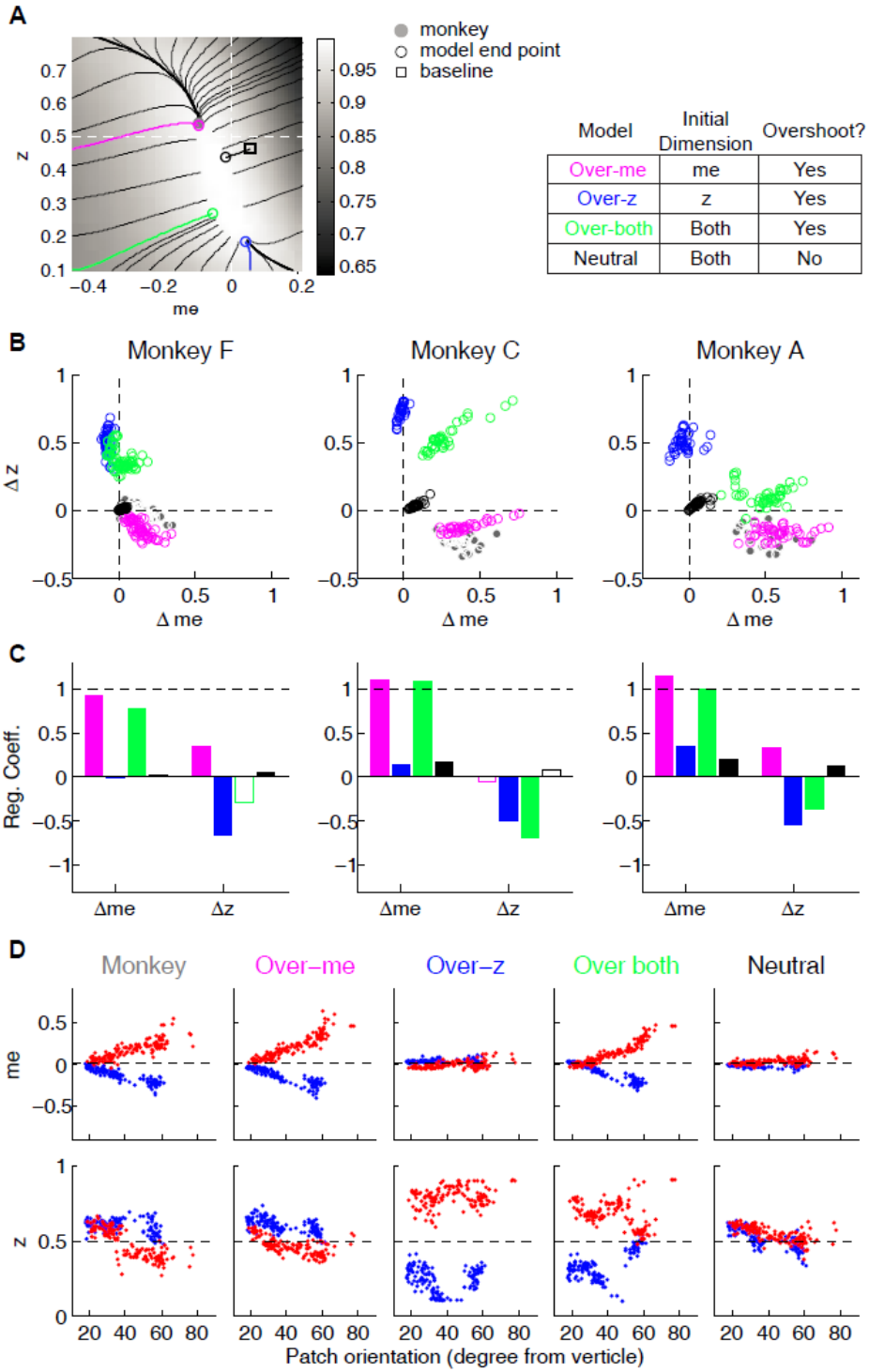
B, Scatterplots of the predicted and actual  $\Delta me$  and  $\Delta z$  between reward contexts. Grey circles here are the same as the black circles in Figure 4C. Colors indicate model identity, as in A.

C, Average regression coefficients between each monkey's  $\Delta me$  (left four bars) and  $\Delta z$  (right four bars) values and predicted values for each of the four models. Filled bars:  $t$ -test,  $p < 0.05$ .

D, Covariation of  $me$  (top) and  $z$  (bottom) with the orientation of the >97% RTrial heatmap patch for monkeys and predictions of the four models. Blue: data from LR-Left blocks, red: data from LR-Right blocks. Data in the "Monkey" column are the same as in Figure 6C and D. Note that predictions of the "over- $me$ " model best matched the monkey data than the other models.

The "over- $me$ " model produced better predictions than the other three alternative models for all three monkeys. Of the four models, only the "over- $me$ " model captured the monkeys' tendency to bias  $me$  toward the large-reward choice (positive  $\Delta me$ ) and bias  $z$  toward the small-reward choice (negative  $\Delta z$ ; Figure 8B). In contrast, the "over- $z$ " model predicted small adjustments in  $me$  and large adjustments in  $z$  favoring the large-reward choice; the "over-both" model predicted relatively large, symmetric  $me$  and  $z$  adjustments favoring the large-reward choice; and the "neutral" model predicted relatively small, symmetric adjustments in both  $me$  and  $z$  favoring the large-reward choice. Accordingly, for each monkey, the predicted and actual values of both  $\Delta me$  and  $\Delta z$  were most strongly positively correlated for predictions from the "over- $me$ " model

compared to the other models (Figure 2.8C). The “over-*me*” model was also the only one of the models we tested that recapitulated the measured relationships between both *me*- and *z*-dependent biases and session-by-session changes in the orientation of the RTrial function (Figure 2.8D). Similar results were observed using RR function (Figure 2.7–figure supplement 1 and Figure 2.8–figure supplement 1). We also examined whether the shape of the reward surface alone can explain the monkeys' bias patterns. We repeated the simulations using randomized starting points, with or without additional noise in each updating step. These simulations could not reproduce the monkeys' bias patterns (data not shown), suggesting that using “over-*me*” starting points is critical for accounting for the monkeys' suboptimal behavior.



**Figure 2.8–figure supplement 1. Predictions of a RR gradient-based model.** Same format as Figure 8 but using gradients based on RR instead of RTrial. The overly-biased starting *me* and *z* values were set as 90% of highest coherence level, and 0.1, respectively, except for the *over-both* model for one monkey C session ( $me = 88\% * \max(\text{coh})$ ,  $z = 0.11$ ) to avoid a local peak in the RR surface. Such local peaks at overly biased *me* and *z* values can divert the gradient-based updating process to even more biased values without ever reaching the monkey's final RR (e.g., the green trace at the bottom left corner in monkey C's LR-Left data in Figure 7–figure supplement 1).

## Discussion

We analyzed the behavior of three monkeys performing a decision task that encouraged the use of both uncertain visual motion evidence and the reward context. All three monkeys made choices that were sensitive to the strength of the sensory evidence and were biased toward the larger-reward choice, which is roughly consistent with previous studies of humans and monkeys performing similar tasks (Maddox and Bohil, 1998; Voss et al., 2004; Diederich and Busemeyer, 2006; Liston and Stone, 2008; Serences, 2008; Feng et al., 2009; Simen et al., 2009; Nomoto et al., 2010; Summerfield and Koechlin, 2010; Teichert and Ferrera, 2010; Gao et al., 2011; Leite and Ratcliff, 2011; Mulder et al., 2012; Wang et al., 2013; White and Poldrack, 2014). However, we also found that these adjustments differed considerably in detail for the three monkeys, in terms of overall magnitude, dependence on perceptual sensitivity and offered rewards, and relationship to RTs. We quantified these effects with a logistic analysis and a commonly used model of decision-making, the drift-diffusion model (DDM), which allowed us to compare the underlying decision-related computations to hypothetical benchmarks that would maximize reward. We found that all three monkeys made reward context-dependent adjustments with two basic components: 1) an over-adjustment of the momentary evidence provided by the sensory stimulus (*me*) in favor of the large-reward

option; and 2) an adjustment to the decision rule that governs the total evidence needed for each choice ( $z$ ), but in the opposite direction (i.e., towards the small-reward option). Similar to some earlier reports of human and monkey performance on somewhat similar tasks, our monkeys did not optimize reward rate (Starns and Ratcliff, 2010 and 2012; Teichert and Ferrera, 2010). Instead, these adjustments tended to provide nearly, but not exactly, maximal reward intake. We proposed a common heuristic strategy based on the monkeys' individual reward functions to account for the idiosyncratic adjustments across monkeys and across sessions within the same monkey.

### ***Considerations for assessing optimality and rationality***

Assessing decision optimality requires a model of the underlying computations. In this study, we chose the DDM for several reasons. First, it provided a parsimonious account of both the choice and RT data (Palmer et al., 2005; Ratcliff et al., 1999). Second, as discussed in more detail below, the DDM and related accumulate-to-bound models have provided useful guidance for identifying neural substrates of the decision process (Roitman and Shadlen, 2002; Ding and Gold, 2010; Ding and Gold, 2012; Hanks et al., 2011; Ratcliff et al., 2003; Rorie et al., 2010; Mulder et al., 2012; Summerfield and Koechlin, 2010; Frank et al., 2015). Third, these models are closely linked to normative theory, including under certain assumptions matching the statistical procedure known as the sequential probability ratio test that can optimally balance the speed and accuracy of uncertain decisions (Barnard, 1946; Wald, 1947; Wald and Wolfowitz, 1948, Edward, 1965). These normative links were central to our ability to use the DDM to relate the monkeys' behavior to different forms of reward optimization. The particular form of DDM that we used produced reasonably good, but not perfect, fits to

the monkeys' data. These results support the utility of the DDM framework but also underscore the fact that we do not yet know the true model, which could impact our optimality assessment.

Assessing optimality also requires an appropriate definition of the optimization goal. In our study, we mainly focused on the goal of maximizing reward rate (per trial or per unit of time). Based on this definition, the monkeys showed suboptimal reward-context-dependent adjustments. It is possible that the monkeys' were optimizing for a different goal, such as accuracy or a competition between reward and accuracy ("COBRA," Maddox and Bohil, 1998). However, the monkeys' behavior was not consistent with optimizing for these goals, either. Specifically, none of these goals would predict optimal  $z$  adjustment that favors the small reward choice: accuracy maximization would require unbiased decisions ( $me=0$  and  $z=0.5$ ), whereas COBRA would require  $z$  values with smaller magnitude (between 0.5 and those predicted for reward maximization alone), but still in the adaptive direction. Therefore, the monkeys' strategies were not consistent with simply maximizing commonly considered reward functions.

Deviations from optimal behavior are often ascribed to a lack of effort or poor learning. However, these explanations seem unlikely to be primary sources of suboptimality in our study. For example, lapse rates, representing the overall ability to attend to and perform the task, were consistently near zero for all three monkeys. Moreover, the monkeys' reward outcomes ( $RTrial$  or  $RR$  with respect to optimal values) did not change systematically with experience but instead stayed close to the optimal values. These results imply that the monkeys understood the task demands and performed consistently well over the course of our study. Suboptimal performance has

also been observed in human subjects, even with explicit instructions about the optimality criteria (Starns and Ratcliff 2010, 2012), suggesting that additional factors need to be considered to understand apparent suboptimality in general forms of decision-making. More importantly, the monkeys made adjustments that were adapted to changes in their idiosyncratic, context-dependent reward functions, which reflected session-specific reward ratios and motion coherences and the monkeys' daily variations of perceptual sensitivity and speed-accuracy trade-offs (Figure 2.6, Figure 2.6–figure supplement 1). Based on these observations, we reasoned that the seemingly sub-optimal behaviors may instead reflect a common, adaptive, rational strategy that aimed to attain good-enough (satisficing) outcomes.

The gradient-based, satisficing model we proposed was based on the considerations discussed below to account for our results. We do not yet know how well this model generalizes to other tasks and conditions, but it exemplifies an additional set of general principles for assessing the rationality of decision-making behavior: goals that are not necessarily optimal but good enough, potential heuristic strategies based on the properties of the utility function, and flexible adaptation to changes in the external and internal conditions.

### ***Assumptions and experimental predictions of the proposed learning strategy***

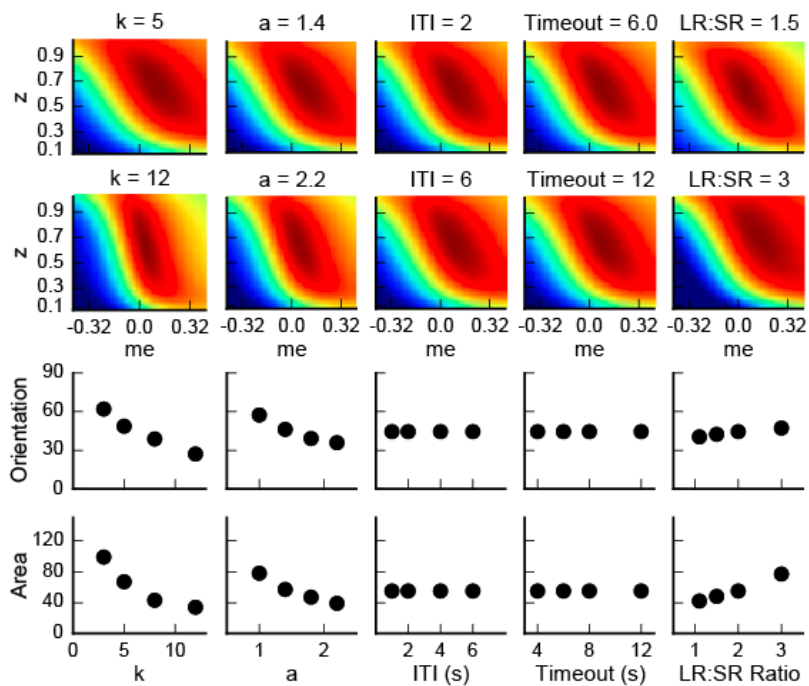
In general, finding rational solutions through trial-and-error or stepwise updates requires a sufficient gradient in the utility function to drive learning (Sutton and Barto, 1998). Our proposed scheme couples a standard gradient-following algorithm with principles that have been used to explain and facilitate decisions with high uncertainties, time pressures, and/or complexity to achieve a satisficing solution (Simon, 1966;

Wierzbicki, 1982; Gigerenzer and Goldstein, 1996; Nosofsky and Palmeri, 1997; Goodrich et al., 1998; Sakawa and Yauchi, 2001; Goldstein and Gigerenzer, 2002; Stirling, 2003; Gigerenzer, 2010; Oh et al., 2016). This scheme complements but differs from a previously proposed satisficing strategy to account for human subjects' suboptimal calibration of the speed-accuracy trade-off via adjustments of the decision bounds of a DDM that favor robust solutions given uncertainties about the inter-trial interval (Zacksenhouse et al., 2010). In contrast, our proposed strategy focuses on reward-biased behaviors for a given speed-accuracy tradeoff and operates on reward per trial, which is, by definition, independent of inter-trial-interval.

Our scheme was based on four key assumptions, as follows. Our first key assumption was that the starting point for gradient following was not the unbiased state (i.e.,  $m_e=0$  and  $z=0.5$ ) but an over-biased state. Notably, in many cases the monkeys could have performed as well or better than they did, in terms of optimizing reward rate, by making unbiased decisions. The fact that none did so prompted our assumption that their session-by-session adjustments tended to reduce, not inflate, biases. Specifically, we assumed that the initial experience of the asymmetric reward prompted an over-reaction to bias choices towards the large-reward alternative. In general, such an initial over-reaction is not uncommon, as other studies have shown excessive, initial biases that are reduced or eliminated with training (Gold et al., 2008; Jones, et al., 2015; Nikolaev et al., 2016). The over-reaction is also rational because the penalty is larger for an under-reaction than for an over-reaction. For example, in the average  $R_{\text{Trial}}$  heatmaps for our task (Figure 2.6A), the gradient dropped faster in the under-biased side than in the over-biased side. This pattern is generally true for tasks with sigmoid-like psychometric functions (for example, the curves in Figure 2.2—figure supplement 2). Our



model further suggests that the nature of this initial reaction, which may be driven by individually tuned features of the reward function that can remain largely consistent even for equal-reward tasks (Figure 2.8–figure supplement 2) and then constrain the endpoints of a gradient-based adjustment process (Figure 2.8), may help account for the extensive individual variability in biases that has been reported for reward-biased perceptual tasks (Voss et al., 2004; Summerfield and Koechlin, 2010; Leite and Ratcliff, 2011; Cicmil et al., 2015)).

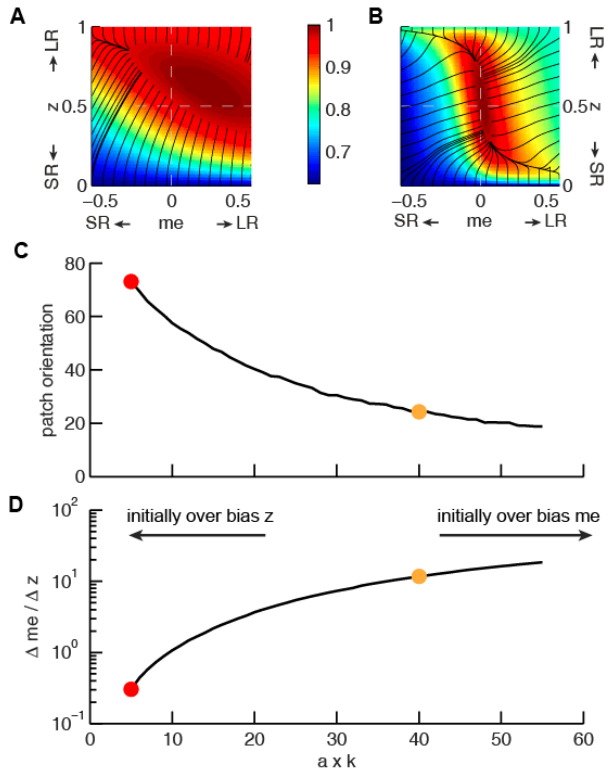


**Figure 2.8–figure supplement 2. Dependence of the orientation and area of the near-optimal RTrial patch on parameters reflecting internal decision process and external task specifications.**

The top two rows show the RTrial heatmaps with two values of a single parameter indicated above, while keeping the other parameters fixed at the baseline values. The third and fourth rows show the estimated orientation (the amount of tilt from vertical, in degrees) and area (in pixels), respectively, of the image patches corresponding to  $\geq 97\%$  of  $R_{\text{Trial}_{\text{max}}}$ . The baseline values of the parameters are:  $a=1.5$ ,  $k=6$ , non-decision times=0.3 sec for both choices,  $ITI=4$  sec,  $Timeout=8$  sec, *large-reward (LR): small-reward (SR) ratio*=2.

The specific form of initial over-reaction in our model, which was based on the gradient asymmetry of the reward function, makes testable predictions. Specifically, our data were most consistent with an initial bias in momentary evidence ( $me$ ), which caused the biggest change in the reward function. However, this gradient asymmetry can change dramatically under different conditions. For example, changes in the subject's cautiousness (i.e., the total bound height parameter,  $a$ ) and perceptual sensitivity ( $k$ ) would result in a steeper gradient in the other dimension (the decision rule, or  $z$ ) of the reward function (Figure 2.8–figure supplement 3). Our model predicts that such a subject would be more prone to an initial bias along that dimension. This prediction can be tested by using speed-accuracy instructions to affect the bound height and different stimulus parameters to change perceptual sensitivity (Palmer et al 2005; Gegenfurtner and Hawken, 1996).

Our second key assumption was that from this initial, over-biased state, the monkeys made adjustments to both the momentary evidence ( $me$ ) and decision rule ( $z$ ) that generally followed the gradient of the reward function. The proposed step-wise adjustments occurred too quickly to be evident in behavior; e.g., the estimated biases were similar for the early and late halves in a block (data not shown). Instead, our primary support for this scheme was that the steady-state biases measured in each session were tightly coupled to the shape of the reward function for that session. It would be interesting to design tasks that might allow for more direct measurements of the updating process itself, for example, by manipulating both the initial biases and relevant reward gradient that might promote a longer adjustment process.



**Figure 2.8–figure supplement 3: The joint effect of DDM model parameters  $a$  (governing the speed-accuracy trade-off) and  $k$  (governing perceptual sensitivity) on the shape of the reward function.**

**(A, B)** Example RTrial functions corresponding to steeper gradients along the  $z$  (panel A, corresponding to the red points in panels C and D) or  $me$  (panel B, corresponding to the orange points in panels C and D) dimension. The gradient lines (black) stop when RTrial  $> 0.97$  of the maximum value. A:  $a=1$ ,  $k=5$ . B:  $a=1$ ,  $k=40$ . Large-reward:small-reward ratio = 2.

**(C)** Orientation of the patch corresponding to  $> 0.97$  maximal RTrial as a function of the product of  $a$  and  $k$ .

**(D)** The ratio of the mean gradients along the  $me$  and  $z$  dimensions as a function of the product of  $a$  and  $k$ . Our model assumes that the initial bias is along the dimension with the steeper gradient according to each monkey's idiosyncratic RTrial function. Note that because  $me$  and  $z$  have different units, the

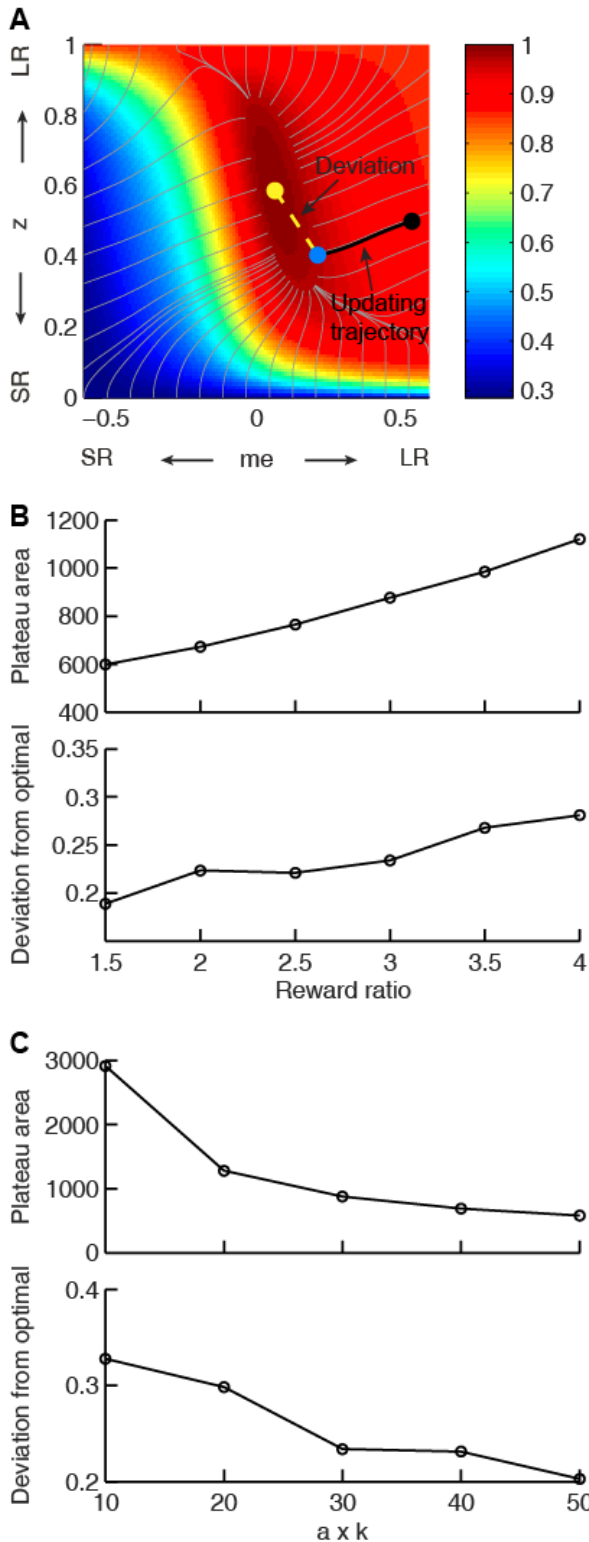
Our third key assumption was that the shallowness of the utility of the function around the peak supported satisficing solutions. Specifically, gradient-based adjustments, particularly those that use rapid updates based on implicit knowledge of the

utility function, may be sensitive only to relatively large gradients. For our task, the gradients were much smaller around the peak, implying that there were large ranges of parameter values that provided such similar outcomes that further adjustments were not used. In principle, it is possible to change the task conditions to test if and how subjects might optimize with respect to steeper functions around the peak. For example, for RTrial, the most effective way to increase the gradient magnitude near the peak (i.e., reducing the area of the dark red patch) is to increase sensory sensitivity ( $k$ ) or cautiousness ( $a$ ; i.e., emphasizing accuracy over speed; Figure 2.8—figure supplement 2). For RR, the gradient can also be enhanced by increasing the time-out penalty. Despite some practical concerns about these manipulations (e.g., increasing time-out penalties can decrease motivation), it would be interesting to study their effects on performance in more detail to understand the conditions under which satisficing or “good enough” strategies are used (Simon, 1956; Simon, 1982).

Our last assumption was that the monkeys terminated adjustments as soon as they reached a good-enough reward outcome. This termination rule produced end points that approximated the monkeys’ behavior reasonably well. Other termination rules are likely to produce similar end points. For example, the learning rate for synaptic weights might decrease as the presynaptic and postsynaptic activities become less variable (Aitchison et al., 2017; Kirkpatrick et al., 2017). In this scheme, learning gradually slows down as the monkey approaches the plateau on the reward surface, which might account for our results.

The satisficing reward gradient-based scheme we propose may further inform appropriate task designs for future studies. For example, our scheme implies that the shape of the reward function near the peak, particularly the steepness of the gradient,

can have a strong impact on how closely a subject comes to the optimal solution for a given set of conditions. Thus, task manipulations that affect the shape of the reward-function peak could, in principle, be used to control whether a study focuses on more- or less-optimal behaviors (Figure 2.8–figure supplement 4). For example, increasing perceptual sensitivity (e.g., via training) and/or decisions that emphasize accuracy over speed (e.g., via instructions) tends to sharpen the peak of the reward function. According to our scheme, this sharpening should promote increasingly optimal decision-making, above and beyond the performance gains associated with increasing accuracy, because the gradient can be followed closer to the peak of the reward function. The shape of the peak is also affected by the reward ratio, such that higher ratios lead to larger plateaus, i.e. shallower gradient, near the peak. This relationship leads to the idea that, all else being equal, a smaller reward ratio may be more suitable for investigating principles of near-optimal behavior, whereas a larger reward ratio may be more suitable for investigating the source and principles of sub-optimal behaviors.



**Figure 2.8–figure supplement 4. Effects of the shape of the reward function on deviations from optimality.**

**(A)** Illustration of our heuristic updating model and measurement of deviation of the end point from optimal. Yellow dot: optimal solution. Gray lines: trajectory for gradient ascent, ending at 0.97 maximal RTrial. Black line: trajectory for updating from the starting point (black dot,  $me=0.54$ ,  $z=0.5$ ), which ended at 0.97 maximal RTrial (blue dot). The deviation of the end point from optimal is measured as the distance from the yellow dot to the blue dot (yellow dashed line). The same starting point and ending criterion were used for data shown in B and C.

**(B)** The area of the 0.97 maximal RTrial plateau and end-point deviation from optimal increase with reward ratio. The product of  $a$  and  $k$  is fixed as 30.

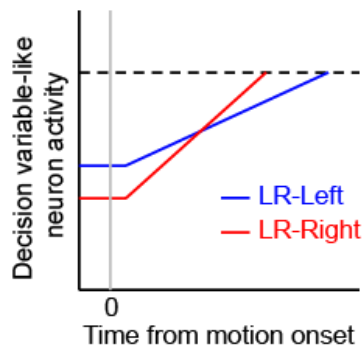
**(C)** The area of the 0.97 maximal RTrial plateau and end-point deviation from optimal decrease with the product of  $a$  and  $k$ . Reward ratio is fixed as 3.

### ***Possible neural mechanisms***

The DDM framework has been used effectively to identify and interpret neural substrates of key computational components of the decision process for symmetric-reward versions of the motion-discrimination task. Our study benefitted from an RT task design that provided a richer set of constraints for inferring characteristics of the underlying decision process than choice data alone (Feng et al., 2009; Nomoto et al., 2010; Teichert and Ferrera, 2010). The monkeys' strategy further provides valuable anchors for future studies of the neural mechanisms underlying decisions that are biased by reward asymmetry, stimulus probability asymmetry, and other task contexts.

For neural correlates of bias terms in the DDM, it is commonly hypothesized that  $m$  adjustments may be implemented as modulation of MT output and/or synaptic weights for the connections between different MT subpopulations and decision areas (Cicmil, et al., 2015). In contrast,  $z$  adjustments may be implemented as context-dependent baseline changes in neural representations of the decision variable and/or context-dependent changes in the rule that determines the final choice (Lo and Wang, 2006; Rao, 2010; Lo et al., 2015; Wei et al., 2015). The manifestation of these adjustments in neural activity that encodes a decision variable may thus differ in its temporal characteristics: a  $m$  adjustment is assumed to modulate the rate of change in neural activity, whereas a  $z$  adjustment does not. However, such a theoretical difference can be challenging to observe, because of the stochasticity in spike generation and, given such stochasticity, practical difficulties in obtaining sufficient data with long decision deliberation times. By adjusting  $m$  and  $z$  in opposite directions, our monkeys' strategies may allow a simpler test to disambiguate neural correlates of  $m$  and  $z$ . Specifically, a neuron or neuronal population that encodes  $m$  may show reward modulation congruent

with its choice preference, whereas a neuron or neuronal population that encodes  $z$  may show reward modulation opposite to its choice preference (Figure 2.8–figure supplement 5). These predictions further suggest that, although it is important to understand if and how human or animal subjects can perform a certain task optimally, for certain systems-level questions, there may be benefits to tailoring task designs to promote sub-optimal strategies in otherwise well-trained subjects.



**Figure 2.8–figure supplement 5. Hypothetical neural activity encoding a reward-biased perceptual decision variable.** The blue and red curves depict rise-to-threshold dynamics in favor of a particular (say, rightward) choice under the two reward contexts, as indicated. Note that when the rightward choice is paired with larger reward: 1) the slope of the ramping process, which corresponds to an adjustment in momentary evidence ( $me$ ), is steeper; and 2) the baseline activity, which corresponds to the decision-rule ( $z$ ) adjustment, is lower.



## **Material and Methods**

### ***Subjects***

We used three rhesus macaques (*Macaca mulatta*), two male and one female, to study behavior on an asymmetric-reward reaction-time random-dot motion discrimination task (Figure 1B, see below). Prior to this study, monkeys F and C had been trained extensively on the equal-reward RT version of the task (Ding and Gold, 2010, 2012b, a). Monkey A had been trained extensively on non-RT dots tasks (Connolly et al., 2009; Bennur and Gold, 2011), followed by >130 sessions of training on the equal-reward RT dots task. All training and experimental procedures were in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the University of Pennsylvania Institutional Animal Care and Use Committee (#804726).

### ***Behavioral task***

Our task (Figure 2.1B) was based on the widely used random-dot motion discrimination task that typically has symmetric rewards (Roitman and Shadlen, 2002; Ding and Gold, 2010). Briefly, a trial started with presentation of a fixation point at the center of a computer screen in front of a monkey. Two choice targets appeared 0.5 s after the monkey acquired fixation. After a delay, the fixation point was dimmed and a random-dot kinematogram (speed: 6 °/s) was shown in a 5° aperture centered on the fixation point. For monkeys F and C, the delay duration was drawn from a truncated exponential distribution with mean=0.7 s, max=2.5 s, min=0.4 s. For monkey A, the delay was set as 0.75 s. The monkey was required to report the perceived global motion direction by making a saccade to the corresponding choice target at a self-determined

time (a 50-ms minimum latency was imposed to discourage fast guesses). The stimulus was immediately turned off when the monkeys' gaze left the fixation window (4, 4, and 3° square windows for monkey F, C, and A, respectively). Correct choices (i.e., saccades to the target congruent with actual motion direction) were rewarded with juice. Error choices were not rewarded and instead penalized with a timeout before the next trial began (timeout duration: 3 s, 0.5-2 s, and 2.5 s, for monkeys F, C, and A, respectively). On each trial, the motion direction was randomly selected toward one of the choice targets along the horizontal axis. The motion strength of the kinematogram was controlled as the fraction of dots moving coherently to one direction (coherence). On each trial, coherence was randomly selected from 0.032, 0.064, 0.128, 0.256, and 0.512 for monkeys F and C, and from 0.128, 0.256, 0.512, and 0.75 for monkey A. In a subset of sessions, coherence levels of 0.064, 0.09, 0.35, and/or 0.6 were also used for monkey A.

We imposed two types of reward context on the basic task. For the "LR-Left" reward context, correct leftward saccades were rewarded with a larger amount of juice than correct rightward saccades. For the "LR-Right" reward context, correct leftward saccades were rewarded with a smaller amount of juice than correct rightward saccades. The large:small reward ratio was on average 1.34, 1.91, and 2.45 for monkeys F, C, and A, respectively. Reward context was alternated between blocks and constant within a block. Block changes were signaled to the monkey with an inter-block interval of 5 s. The reward context for the current block was signaled to the monkey in two ways: 1) in the first trial after a block change, the two choice targets were presented in blue and green colors, for small and large rewards, respectively (this trial was not included for analysis); and 2) only the highest coherence level (near 100% accuracy)

was used for the first two trials after a block change to ensure that the monkey physically experienced the difference in reward outcome for the two choices. For the rest of the block, choice targets were presented in the same color and motion directions and coherence levels were randomly interleaved.

We only included sessions in which there are more than 200 trials, more than 8 coherences and more than 8 trials for each coherence, motion direction and reward context (61, 37 and 43 sessions for monkey F, C and A, respectively).

### ***Basic characterization of behavioral performance***

Eye position was monitored using a video-based system (ASL) sampled at 240 Hz. RT was measured as the time from stimulus onset to saccade onset, the latter identified offline with respect to velocity ( $> 40^\circ/\text{s}$ ) and acceleration ( $> 8000^\circ/\text{s}^2$ ). Performance was quantified with psychometric and chronometric functions (Figure 2 and Figure 3), which describe the relationship of motion strength (signed coherence,  $Coh$ , which was the proportion of the dots moving in the same direction, positive for rightward motion, negative for leftward motion) with choice and RT, respectively. Psychometric functions were fitted to a logistic function (Equation (1)), in which  $\lambda$  is the error rate, or lapse rate, independent of the motion information;  $\alpha_0$  and  $(\alpha_0 + \alpha_{rew})$  are the bias terms, which measures the coherence at which the performance was at chance level in the LR-Right and LR-Left reward contexts, respectively.  $\beta_0$  and  $(\beta_0 + \beta_{rew})$  are the perceptual sensitivities in the LR-Right and LR-Left reward contexts, respectively.

$$P_{rightward\ choice} = \lambda + (1 - 2\lambda) \times \frac{1}{e^{-Sensitivity(Coh-Bias)}} \quad (1)$$

### ***Reward-biased drift-diffusion model***

To infer the computational strategies employed by the monkeys, we adopted the widely used accumulation-to-bound framework, the drift-diffusion model (DDM; Figure 1A). In the standard DDM, motion evidence is modeled as a random variable following a Gaussian distribution with a mean linearly proportional to the signed coherence and a fixed variance. The decision variable (DV) is modeled as temporal accumulation (integral) of the evidence, drifting between two decision bounds. Once the DV crosses a bound, evidence accumulation is terminated, the identity of the decision is determined by which bound is crossed, and the decision time is determined by the accumulation time. RT is modeled as the sum of decision time and saccade-specific non-decision times, the latter accounting for the contributions of evidence-independent sensory and motor processes.

To model the observed influences of motion stimulus and reward context on monkeys' choice and RT behavior, we introduced two reward context-dependent terms:  $z$  specifies the relative bound heights for the two choices and  $me$  specifies the equivalent momentary evidence that is added to the motion evidence at each accumulating step. Thus, for each reward context, six parameters were used to specify the decision performance:  $a$ : total bound height;  $k$ : proportional scaling factor converting evidence to the drift rate;  $t_0$  and  $t_1$ : non-decision times for leftward and rightward choices, respectively; and  $z$  and  $me$ . Similar approaches have been used in studies of human and animal decision making under unequal payoff structure and/or prior probabilities (Voss et al., 2004; Bogacz et al., 2006; Diederich and Busemeyer, 2006; Summerfield and Koechlin, 2010; Hanks et al., 2011; Mulder et al., 2012).

To fit the monkeys' data, we implemented hierarchical DDM fitting using an open-source package in Python, which performs Bayesian estimates of DDM parameters

based on single-trial RTs (Wiecki et al., 2013). This method assumes that parameters from individual sessions are samples from a group distribution. The initial prior distribution of a given parameter is determined from previous reports of human perceptual performance and is generally consistent with monkey performance on equal reward motion discrimination tasks (Ding and Gold, 2010; Matzke and Wagenmakers, 2009). The posterior distributions of the session- and group-level parameters are estimated with Markov chain Monte Carlo sampling. The HDDM was fit to each monkey separately.

For each dataset, we performed 5 chains of sampling with a minimum of 10000 total samples (range: 10000-20000; burn-in: 5000 samples) and inspected the trace, autocorrelation and marginal posterior histogram of the group-level parameters to detect signs of poor convergence. To ensure similar level of convergence across models, we computed the Gelman-Rubin statistic ( $R\text{-hat}$ ) and only accepted fits with  $R\text{-hat} < 1.01$ . To assess whether reward context modulation of both  $z$  and  $me$  was necessary to account for monkeys' behavioral data, we compared fitting performance between the model with both terms ("full") and reduced models with only one term ("z-only" and "me-only"). Model selection was based on the deviance information criterion (DIC), with a smaller DIC value indicating a preferred model. Because DIC tends to favor more complex models, we bootstrapped the expected  $\Delta$ DIC values, assuming the reduced models were the ground truth, using trial-matched simulations. For each session, we generated simulated data using the DDM, with single-session parameters fitted by  $me$ -only or  $z$ -only HDDM models and with the number of trials for each direction  $\times$  coherence  $\times$  reward context combination matched to the monkey's data for that session.

These simulated data were then re-fitted by all three models to estimate the predicted  $\Delta$ DIC, assuming the reduced model as the generative model.

To test an alternative model, we also fitted monkeys' data to a DDM with collapsing bounds (Zylberberg et al., 2016). This DDM was constructed as the expected first-stopping-time distribution given a set of parameters, using the PyMC module (version 2.3.6) in Python (version 3.5.2). The three model variants, "full", "me-only" and "z-only", and their associated parameters were the same as in HDDM, except that the total bound distance decreases with time. The distance between the two choice bounds was set as  $a/(1 + e^{\beta(t-d)})$ , where  $a$  is the initial bound distance,  $\beta$  determines the rate of collapsing, and  $d$  determines the onset of the collapse. Fitting was performed by computing the maximum *a posteriori* estimates, followed by Markov chain Monte Carlo sampling, of DDM parameters given the experimental RT data.

### **Sequential analysis**

To examine possible sequential choice effects, for each monkey and session we fitted the choice data to three logistic functions. Each function was in the same form as equation (1) but with one of four possible additional terms describing a sequential effect based on whether the previous trial was correct or not, and whether the previous trial was to the large or small reward target. The sequential effect was assessed via a likelihood-ratio test for  $H_0$ : the sequential term in Eq. (2)=0,  $p < 0.05$

$$P_{\text{rightward choice}} = \lambda + (1 - 2\lambda) \times \frac{1}{e^{-\text{Sensitivity}(\text{Coh} - (\text{Bias} + \text{Bias}_{\text{seq}}))}} \quad (2)$$

$Bias_{seq}$  was determined using indicator variables for the given sequential effect and the reward context (e.g., LR-Right context, previous correct LR choice):  $I_{seq} \times I_{rew} \times \alpha_{seq}$ , where

$I_{rew} = +/-1$  for LR-Right / LR-Left reward contexts.

$I_{seq} = I_{prevLR-prevCorrect}$ ,  $I_{prevLR-prevError}$ ,  $I_{prevSR-prevCorrect}$ , and  $I_{prevSR-prevError}$  for the 4 types of sequential effects (note that there were not enough trials to compute previous error SR choice).

### **Optimality analysis**

To examine the level of optimality of the monkeys' performance, we focused on two reward functions: reward rate (RR, defined as the average reward per second) and reward per trial (RTrial, defined as the average reward per trial) for a given reward context for each session. To estimate the reward functions in relation to  $me$  and  $z$  adjustments for a given reward context, we numerically obtained choice and RT values for different combinations of  $z$  (ranging from 0 to 1) and  $me$  (ranging from -0.6 to 0.6 coherence unless otherwise specified), given  $a$ ,  $k$  and non-decision time values fitted by the full model. We then calculated RR and RTrial, using trial-matched parameters, including the actual ITI, timeout, and large:small reward ratio.  $RR_{max}$  and  $RTrial_{max}$  were identified as the maximal values given the sampled  $me$ - $z$  combinations, using 1000 trials for each coherence  $\times$  direction condition. Optimal  $me$  and  $z$  adjustments were defined as the  $me$  and  $z$  values corresponding to  $RR_{max}$  or  $RTrial_{max}$ .  $RR_{predict}$  and  $RTrial_{predict}$  were calculated with the fitted  $me$  and  $z$  values in the full model.

## References

- Aitchison L, Pouget A, Latham P (2017) Probabilistic Synapses. arXiv:1410.1029.
- Ashby FG (1983) A Biased Random-Walk Model for 2 Choice Reaction-Times. *Journal of Mathematical Psychology* 27:277-297.
- Barnard GA (1946) Sequential Tests in Industrial Statistics. *J Roy Stat Soc B* 8:1-26.
- Bennur S, Gold JI (2011) Distinct representations of a perceptual decision and the associated oculomotor plan in the monkey lateral intraparietal area. *J Neurosci* 31:913-921.
- Blank H, Biele G, Heekeren HR, Philiastides MG (2013) Temporal characteristics of the influence of punishment on perceptual decision making in the human brain. *J Neurosci* 33:3939-3952.
- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113:700-765.
- Churchland AK, Kiani R, Shadlen MN (2008) Decision-making with multiple alternatives. *Nat Neurosci* 11:693-702.
- Churchland AK, Kiani R, Chaudhuri R, Wang XJ, Pouget A, Shadlen MN (2011) Variance as a signature of neural computations during decision making. *Neuron* 69:818-831.
- Cicmil N, Cumming BG, Parker AJ, Krug K (2015) Reward modulates the effect of visual cortical microstimulation on perceptual decisions. *eLife* 4.
- Connolly PM, Bennur S, Gold JI (2009) Correlates of perceptual learning in an oculomotor decision variable. *J Neurosci* 29:2136-2150.



- Diederich A, Busemeyer JR (2006) Modeling the effects of payoff on response bias in a perceptual discrimination task: bound-change, drift-rate-change, or two-stage-processing hypothesis. *Percept Psychophys* 68:194-207.
- Ding L (2015) Distinct dynamics of ramping activity in the frontal cortex and caudate nucleus in monkeys. *J Neurophysiol* 114:1850-1861.
- Ding L, Gold JI (2010) Caudate encodes multiple computations for perceptual decisions. *J Neurosci* 30:15747-15759.
- Ding L, Gold JI (2012a) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cereb Cortex* 22:1052-1067.
- Ding L, Gold JI (2012b) Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. *Neuron* 75:865-874.
- Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The cost of accumulating evidence in perceptual decision making. *J Neurosci* 32:3612-3628.
- Edwards W (1965) Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *J Mathematical Psychology* 2(2): 312-329.
- Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, Cavanagh JF and Badre D (2015) fMRI and EEG Predictors of Dynamic Decision Parameters during Human Reinforcement Learning. *J Neurosci* 35: 484-494.
- Feng S, Holmes P, Rorie A, Newsome WT (2009) Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS computational biology* 5:e1000284.

- Gao JA, Tortell R, McClelland JL (2011) Dynamic Integration of Reward and Stimulus Information in Perceptual Decision-Making. *PLoS One* 6.
- Gegenfurtner KR, Hawken MJ (1996) Interaction of motion and color in the visual pathways. *Trends Neurosci* 19:394-401.
- Gigerenzer G (2010) Moral satisficing: rethinking moral behavior as bounded rationality. *Topics in cognitive science* 2:528-554.
- Gigerenzer G, Goldstein DG (1996) Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103:650-669.
- Gigerenzer G, Gaissmaier W (2011) Heuristic Decision Making. *Annual Review of Psychology* 62:451-482.
- Gold JI, Shadlen MN (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* 36:299-308.
- Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535-574.
- Goldfarb S, Leonard NE, Simen P, Caicedo-Nunez CH, Holmes P (2014) A comparative study of drift diffusion and linear ballistic accumulator models in a reward maximization perceptual choice task. *Front Neurosci* 8:148.
- Goldstein DG, Gigerenzer G (2002) Models of ecological rationality: the recognition heuristic. *Psychol Rev* 109:75-90.
- Goodrich MA, Stirling WC, Frost RL (1998) A theory of satisficing decisions and control. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 28:763-779.
- Hanks TD, Ditterich J, Shadlen MN (2006) Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nat Neurosci* 9:682-689.

- Hanks TD, Mazurek ME, Kiani R, Hopp E, Shadlen MN (2011) Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J Neurosci* 31:6339-6352.
- Jones PR, Moore DR, Shub DE, Amitay S (2015) The Role of Response Bias in Perceptual Learning. *J Exp Psychol Learn Mem Cogn*. 41:1456–1470.
- Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324:759-764.
- Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D, Hadsell R (2017) Overcoming catastrophic forgetting in neural networks. *Proc Natl Acad Sci U S A* 114(13):3521-3526.
- Klein SA (2001) Measuring, estimating, and understanding the psychometric function: a commentary. *Percept Psychophys* 63:1421-1455.
- Horwitz GD, Newsome WT (2001) Target selection for saccadic eye movements: Direction selective visual responses in the superior colliculus. *J Neurophysiol*. 86:2527-2542
- Krajbich I, Armel C, Rangel A (2010) Visual fixations and the computation and comparison of value in simple choice. *Nat Neurosci* 13:1292-1298.
- Latimer KW, Yates JL, Meister ML, Huk AC, Pillow JW (2015) Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science* 349:184-187.
- Leite FP, Ratcliff R (2011) What cognitive processes drive response biases? A diffusion model analysis. *Judgm Decis Mak* 6:651-687.

- Liston DB, Stone LS (2008) Effects of prior information and reward on oculomotor and perceptual choices. *J Neurosci* 28:13866-13875.
- Lo CC, Wang XJ (2006) Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci* 9:956-963.
- Lo CC, Wang CT, Wang XJ (2015) Speed-accuracy tradeoff by a control signal with balanced excitation and inhibition. *J Neurophysiol* 114:650-661.
- Maddox WT, Bohil CJ (1998) Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of experimental psychology Learning, memory, and cognition* 24:1459-1482.
- Matzke D, Wagenmakers EJ (2009) Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychon Bull Rev* 16:798-817.
- Milosavljevic M, Malmaud J, Huth A, Koch C, Rangel A. (2001) The Drift Diffusion Model can account for the accuracy and reaction times of value-based choice under high and low time pressure. *Judgment and Decision Making* 5:437-449.
- Mulder MJ, Wagenmakers EJ, Ratcliff R, Boekel W, Forstmann BU (2012) Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J Neurosci* 32:2335-2343.
- Newsome WT, Britten KH, Movshon JA (1989) Neuronal correlates of a perceptual decision. *Nature* 341:52-54.
- Nikolaev AR, Gepshtein S, van Leeuwen C (2016) Intermittent regime of brain activity at the early, bias-guided stage of perceptual learning. *J Vis.* 16(14):11.

- Nomoto K, Schultz W, Watanabe T, Sakagami M (2010) Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. *J Neurosci* 30:10692-10702.
- Nosofsky RM, Palmeri TJ (1997) An exemplar-based random walk model of speeded classification. *Psychol Rev* 104:266-300.
- Oh H, Beck JM, Zhu P, Sommer MA, Ferrari S, Egnér T (2016) Satisficing in split-second decision making is characterized by strategic cue discounting. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42:1937-1956.
- Palmer J, Huk AC, Shadlen MN (2005) The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of vision* 5:376-404.
- Rao RP (2010) Decision making under uncertainty: a neural model based on partially observable markov decision processes. *Frontiers in computational neuroscience* 4:146.
- Ratcliff R (1978) Theory of Memory Retrieval. *Psychological Review* 85:59-108.
- Ratcliff R (1985) Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychol Rev.* 92(2):212-25.
- Ratcliff R, Tuerlinckx F (2002) Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychon B Rev* 9:438-481.
- Ratcliff R, Cherian A, Segraves M (2003) A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *J Neurophysiol* 90:1392-1407.
- Ratcliff R, Smith PL (2004) A comparison of sequential sampling models for two-choice reaction time. *Psychol Rev* 111:333-367.

- Ratcliff R, Van Zandt T, McKoon G (1999) Connectionist and diffusion models of reaction time. *Psychological Review* 106:261-300.
- Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22:9475-9489.
- Rorie AE, Gao J, McClelland JL, Newsome WT (2010) Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS One*. 5:e9308.
- Sakawa M, Yauchi K (2001) An interactive fuzzy satisficing method for multiobjective nonconvex programming problems with fuzzy numbers through coevolutionary genetic algorithms. *IEEE transactions on systems, man, and cybernetics Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* 31:459-467.
- Serences JT (2008) Value-based modulations in human visual cortex. *Neuron* 60:1169-1181.
- Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol*. 86:1916-36.
- Shadlen MN, Shohamy D (2016) Decision Making and Sequential Sampling from Memory. *Neuron*. 90:927-939.
- Simen P, Contreras D, Buck C, Hu P, Holmes P, Cohen JD (2009) Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *Journal of experimental psychology Human perception and performance* 35:1865-1897.

- Simon HA (1956) Rational choice and the structure of the environment. *Psychol Rev* 63:129-138.
- Simon HA (1966) Theories of Decision-Making in Economics and Behavioural Science. In: *Surveys of Economic Theory: Resource Allocation*, pp 1-28. London: Palgrave Macmillan UK.
- Simon HA (1982) *Models of bounded rationality*. Cambridge, Mass.: MIT Press.
- Smith PL, Ratcliff R (2004) Psychology and neurobiology of simple decisions. *Trends in Neurosciences* 27:161-168.
- Starns JJ, Ratcliff R (2010) The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychol Aging* 25(2):377-90.
- Starns JJ, Ratcliff R (2012) Age-related differences in diffusion model boundary optimality with both trial-limited and time-limited tasks. *Psychon Bull Rev*. 19(1):139-45.
- Stirling WC (2003) *Satisficing games and decision making : with applications to engineering and computer science*. Cambridge, England ; New York: Cambridge University Press.
- Summerfield C, Koechlin E (2010) Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. *Frontiers in Human Neuroscience* 4.
- Sutton RS, Barto A (1998) *Reinforcement learning: An introduction*. Cambridge, Massachusetts, USA; MIT Press.
- Teichert T, Ferrera VP (2010) Suboptimal integration of reward magnitude and prior reward likelihood in categorical decisions by monkeys. *Front Neurosci* 4:186.
- Thura D, Beauregard-Racine J, Fradet CW, Cisek P (2012) Decision making by urgency gating: theory and experimental support. *J Neurophysiol* 108:2912-2930.

- Vandekerckhove J, Tuerlinckx F (2007) Fitting the Ratcliff diffusion model to experimental data. *Psychon B Rev* 14:1011-1026.
- Voss A, Rothermund K, Voss J (2004) Interpreting the parameters of the diffusion model: an empirical validation. *Mem Cognit* 32:1206-1220.
- Wald A (1947) *Sequential analysis*. New York: Wiley.
- Wald A, Wolfowitz J (1948) Optimum Character of the Sequential Probability Ratio Test. *Ann Math Stat* 19:326-339.
- Wang AY, Miura K, Uchida N (2013) The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nat Neurosci* 16:639-647.
- Wei W, Rubin JE, Wang XJ (2015) Role of the indirect pathway of the basal ganglia in perceptual decision making. *J Neurosci* 35:4052-4064.
- White CN, Poldrack RA (2014) Decomposing bias in different types of simple decisions. *Journal of experimental psychology Learning, memory, and cognition* 40:385-398.
- Wiecki TV, Sofer I, Frank MJ (2013) HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics* 7:14.
- Wierzbicki AP (1982) A mathematical basis for satisficing decision making. *Mathematical Modelling* 3:391-405.
- Zacksenhouse M, Bogacz R, Holmes P (2010) Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology* 54:230-246.
- Zylberberg A, Fetsch CR, Shadlen MN (2016) The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *eLife* 5.



# CHAPTER 3: NEURAL REPRESENTATION OF SENSORY AND REWARD INFORMATION IN THE CAUDATE NUCLEUS IN REWARD-BIASED PERCEPTUAL DECISION-MAKING

Yunshu Fan, Takahiro Doi, Joshua I. Gold, Long Ding

Part of this chapter is from a manuscript on BioRxiv: Doi T, Fan Y, Gold JI, Ding L (2019) The caudate nucleus controls coordinated patterns of adaptive, context-dependent adjustments to complex decisions. doi: <https://doi.org/10.1101/568733>

## Introduction

Decision-making is a complex process in which both human and animals have to combine different sources of information, such as noisy sensory evidence and preference for a certain option, in appropriate ways to obtain the outcome that the satisfies the decision-maker. Previous studies have provided many insights into the kinds of computations underlying such adaptive decision-making process, including the ones I described in Chapter 2 (Maddox and Bohil, 1998; Voss et al., 2004; Diederich and Busemeyer, 2006; Whiteley and Sahani, 2008; Liston and Stone, 2008; Feng et al., 2009; Summerfield and Koechlin, 2010; Gao et al., 2011; Leite and Ratcliff, 2011; Mulder et al., 2012; Fan et al., 2018; Waiblinger et al., 2019). However, it remains unclear where and how these computations are implemented in the brain.

A prime candidate for mediating these computations is the basal ganglia pathway, which has been a focus of many modeling studies (Redgrave et al., 1999; Bogacz and Gurney, 2007; Kable and Glimcher, 2009; Rao, 2010; Ratcliff and Frank,

2012; Summerfield and Tsetsos, 2012; Ding and Gold, 2013; Hikosaka et al., 2014).

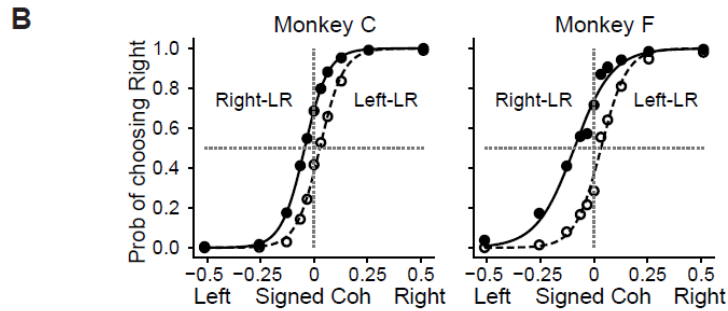
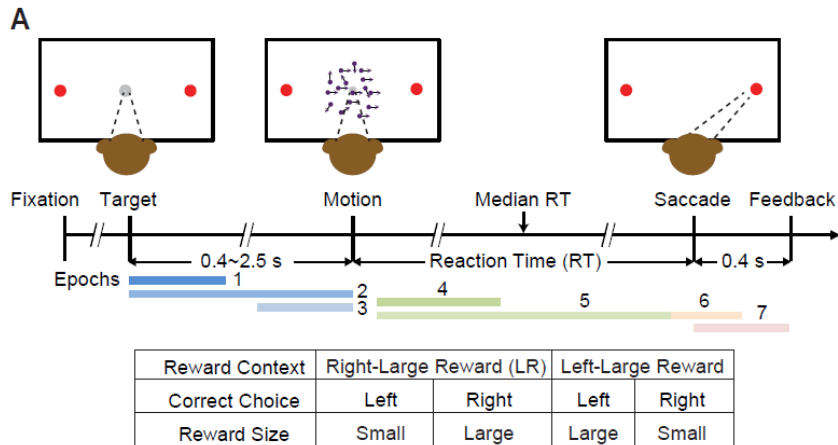
This pathway is known to make separate contributions to perceptual decisions based on the interpretation of uncertain sensory evidence and value-based decisions that select among outcome options (Hikosaka et al., 1989; Nakamura and Hikosaka, 2006; Samejima and Doya, 2007; Lau and Glimcher, 2008; Kimchi and Laubach, 2009; Ding and Gold, 2010, 2012a; Cai et al., 2011; Cavanagh et al., 2011; Seo et al., 2012; Tachibana and Hikosaka, 2012; Tai et al., 2012; Kim and Hikosaka, 2013; Santacruz et al., 2017; Wang et al., 2018; Yartsev et al., 2018). However, its role in combining those different sources of information remains speculative.

To begin to investigate this problem, we performed single-unit extracellular recordings in the caudate nucleus, the input station in the basal ganglia, while monkeys were performing the reward-biased visual motion discrimination task that we used in Chapter 2. In this chapter, I will focus on task- and decision-related information carried by the caudate nucleus before, during and after decision, and how they contribute to the reward-biased decision behavior.

## **Results**

We trained two monkeys to perform the same random-dot visual motion direction discrimination task as described in Chapter 2 (Figure 3.1 A). Both monkeys showed similar reward-biased behaviors as described in Chapter 2 (Figure 3.1 B). While the monkeys were doing the task, we recorded extracellularly from 142 well-isolated units in the caudate nucleus. For data analyses, we divided the task into 7 over-lapping epochs (indicated by the colored bars below the task timeline): 3 epochs before the sensory

stimulus presentation (blue), 2 epochs during decision-making (green) and 2 post-decision epochs (orange and red).



**Figure 3.1 Monkeys showed biases toward choices associated with large reward.**

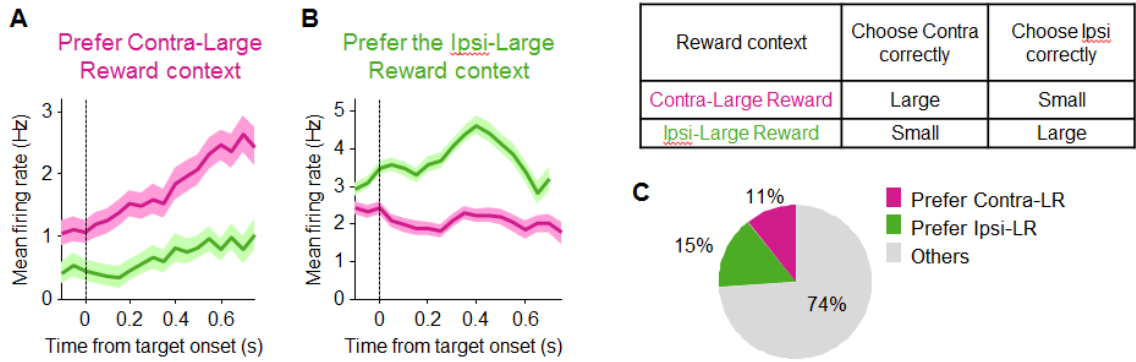
(A) Task design and timeline. Monkeys reported the perceived motion direction with saccades to one of the two choice targets. The motion stimulus was turned off upon detection of saccade. Correct trials were rewarded based on the reward context. Error trials were not rewarded. The color bars in the timeline indicate epoch definitions for the regression analysis of neural firing rates in Eq. 1.

(B) Average choice behavior of two monkeys ( $n = 17,493$  trials from 38 sessions for monkey C, 29,599 trials from 79 sessions for monkey F). Filled and open circles: data from the two reward contexts. Lines: logistic fits.

***Before sensory stimulus presentation, caudate neurons represent reward-context information.***

The behavioral task we used was setup so that the reward context (i.e., which target is associated with large/small reward) was cued by the colors of the two choice targets (blue/green indicated large/small reward) before the starting of a block. Once the block started, the color cues were removed and two identical red choice targets appeared. This protocol required the monkeys to remember the reward context of the block in order to combine this information appropriately with the sensory information presented in each trial. Within each block, the reward context remained the same from trial to trial, which implies that the information about the reward context should be present even before the onset of visual stimulus.

We found that 26% of the 142 caudate neurons that we recorded showed selective preference for one of the reward contexts (beta coefficient of the reward-context regressor being significant in Eq. 1, t-test,  $p < 0.05$ ), including 11% that were more active in the block in which the contralateral correct choice was paired with large reward (such as the example neuron in Figure 3.2 A), and 15% that were more active in the block where the ipsilateral correct choice is paired with large reward (such as the example neuron in Figure 3.2 B). The proportion of reward-context modulation is higher than chance level, which is 5%.



**Figure 3.2 Reward context representations before visual stimulus onset.**

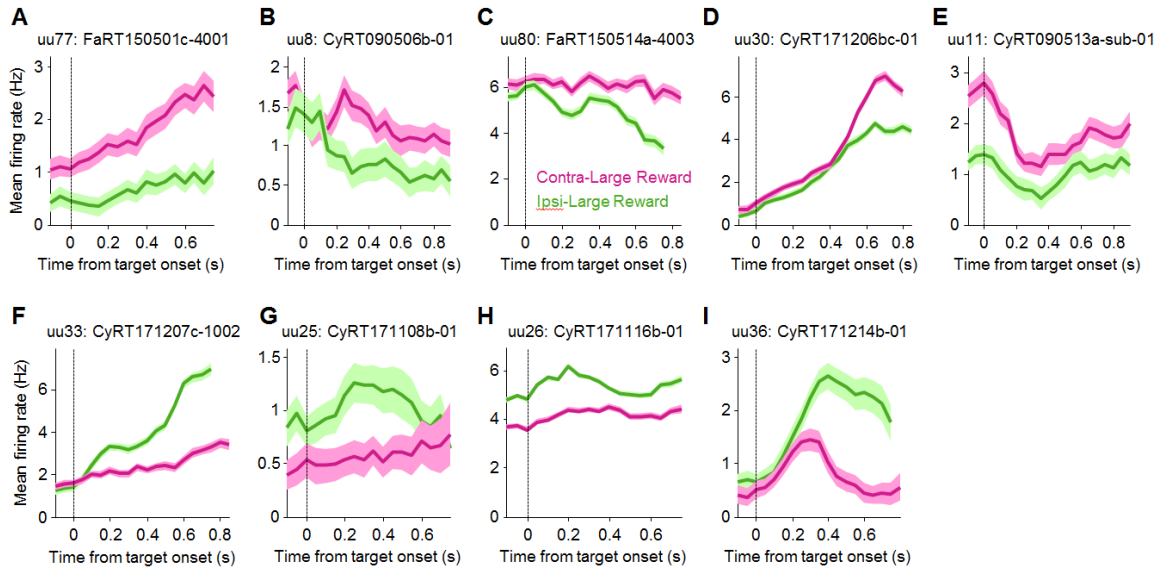
**(A, B)** Activity of two example units before visual stimulus onset. Lines: average firing rate of all trials belonging to each of the two reward contexts. Firing rates were computed using a 200 ms running window (50-ms steps). Ribbon: standard error. Colors: reward context. Note that the neuron in (A) showed preference for the reward context where contralateral target is paired with large reward; the neuron in (B) showed preference for the other reward context.

**(C)** Fraction of neurons representing the two reward contexts or not carrying reward context information. Total number of neurons: 142. Reward context-representing neurons were identified as having significant coefficient ( $t$ -test;  $p < 0.05$ ) for the reward context regressor in the linear regression (Eq.1 in Materials and Methods).

Reward context-dependent pre-decision activity has been reported previously (Lauwereyns et al., 2002; Ding and Hikosaka, 2006). In the study by Ding and Hikosaka, monkeys were asked to make either a contralateral or an ipsilateral saccade toward the previously cued position (memory guided saccade (MGS) task). They used the same block-wise design of asymmetric-reward paradigm as in our task. They found that ~30% of caudate neurons recorded showed preference for one of the two reward contexts, and there were similar proportions of neurons preferring each reward contexts. These findings are consistent with ours. Lauwereyns and colleagues used the asymmetric-reward design in a visually guided saccade (VGS) task, in which the monkey simply needed to look at the target located either on the contralateral or the ipsilateral side. The

VGS task is mentally less challenging than the MGS task because the monkeys do not need to remember the saccade target location. Lauwereyns and colleague found a higher proportion of neurons showing pre-decision reward context modulation (76%) and most of those neurons preferred the context in which the contralateral target is paired with reward (they used reward/no reward, rather than large/small reward). The differences in percentage of reward context-modulated neurons reported by these three studies might due to differences in recording locations and/or sampling bias: for example, neurons active in the VGS task and are modulated by reward context might not be active in the MGS task or in our task. It is also possible that when a task is more cognitively demanding, the proportion of neurons representing the two reward-location associations might be more balanced.

In the two previous studies, the neural activity representing reward context information tended to ramp up until the saccade location cue appeared. We saw similar ramping activity in only 33% neurons based on visual inspection (e.g. Figure 3.2 A and Figure 3.2-figure supplement 1 A and F), but we also saw other patterns (e.g. Figure 3.2 B and Figure 3.2-figure supplement 1 B-E and G-I). A difference between our study and theirs is that, in their studies, the pre-cue period was fixed, so that the monkey could predict when the cue would be turned on, whereas in our study, that period was varied. This might contribute to a lack of consistent ramping pattern.



**Figure 3.2-figure supplement 1. Diverse temporal dynamics of reward context representations before visual stimulus onset.**

(A) - (I) Activity of example units before visual stimulus onset. Same format as (A) and (B) in Figure 3.2.

Top and bottom rows are neurons representing the two reward contexts. Note that neurons in (A) and (F) showed the classical ramping pattern. Neurons in other panels did not show ramping pattern.

In an earlier study using the equal-reward version of our motion-direction discrimination task, caudate neural activity before stimulus onset was shown to correlate with the monkeys' bias towards a specific choice, in both correct and error trials (Ding and Gold, 2010). The correlation was stronger in low coherence trials, because those were the trials where pre-decision bias was less influenced by sensory information. Their results suggested that caudate might encode a choice-bias signal for the monkeys' upcoming decision. We examined whether the choice-bias signal also existed in the caudate pre-stimulus activity in our task using multiple linear regression (Eq.1) in low coherence correct and error trials separately. We found only 1 neuron whose pre-stimulus activity showed significant choice modulation in both correct and error trials.

This suggests that when the monkeys' bias is more reward-driven, caudate pre-stimulus activity appears to represent the reward context, rather than choice bias.

To summarize, caudate neurons can represent reward context information in some neurons before the onset of the sensory stimulus. This reward context information can be used to develop a bias towards the larger reward option, and bias the upcoming decision process, similar to the bias in the starting point of evidence accumulation. We found that the proportions of neurons representing each reward contexts were similar. In the basal ganglia, there are direct and indirect pathways that drive and suppress basal ganglia output, respectively (Purves, 2001). In the context of our results, it is possible that neurons representing the two reward contexts belong to the two separate pathways. Consequently, at the output station of basal ganglia, such as the substantia nigra pars reticulata (SNr) or the Globus Pallidus internal (GPi), the two pathways form a unidirectional bias to increase or decrease the starting point of evidence accumulation in the two reward contexts. Alternatively, if there are two accumulators for evidence supporting each choice, each of the accumulators could receive the information only from neurons representing one of the reward contexts.

***During motion-viewing, caudate neurons represent choice, coherence, reward size and reward context.***

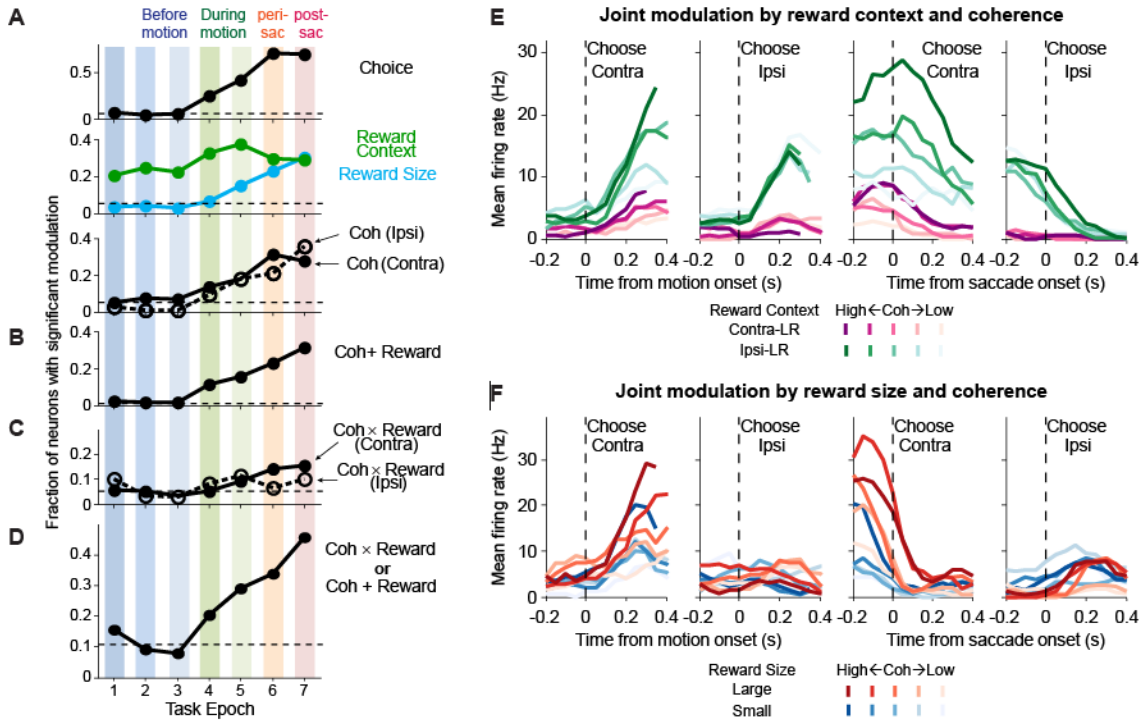
During the motion-viewing period, the monkeys were receiving sensory information about motion direction and strength and they combined that with the reward bias established by the reward context to form their decisions. Because we cued the monkey about the reward context in each block, as a decision is formed, the size of the reward associated with the choice would also be known. We assessed whether a neuron



carries information about the motion strength (i.e., coherence), the reward context, the monkey's choice and the reward size associated with the choice ("reward size" for short), using a multiple linear regression (Eq. 2).

We found that, during this period, the information about reward context was still represented in a significant proportion of caudate neurons (second row in Figure 3.3 A: the proportions of neurons with significant reward context-modulation (green line and circle) during the two motion-viewing epochs were above chance level (dashed line)). Reward context-modulation in some neurons was so obvious that one can deduce reward context change simply from the raw raster plot (two example neurons are shown in Figure 3.3-figure supplement 1).

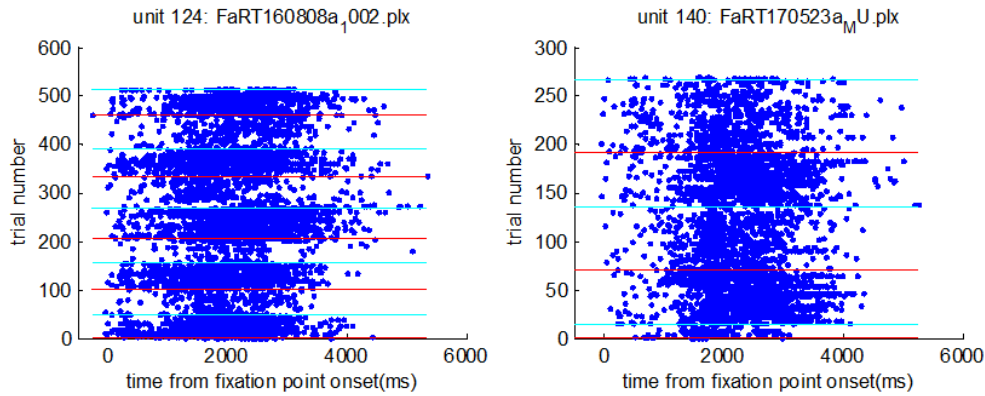
Representation about the motion coherence, choice and reward size information also emerged during this period (Figure 3.3 A: the proportions of neurons with significant modulation by coherence (third row), choice (first row) and reward size (blue line in the second row) during the two motion-viewing epochs were above chance level (dashed line)). The information was represented in more neurons in the later decision epoch (light green) than in the early decision epoch (dark green), consistent with the accumulation of sensory evidence and gradual formation of a decision (Roitman and Shadlen, 2002; Kiani et al., 2008; Yartsev et al., 2018).



**Figure 3.3** Caudate activity reflected motion strength, reward context, choice, and the expected reward size associated with the choice.

(A-D) Fractions of neurons showing significant coefficients for task-related regressors in the seven task epochs defined in Figure 3.1 A (see Eq. 2 for the formulation of regression). Horizontal dashed lines: chance levels. Coh: activity with non-zero coefficients for unsigned coherence values. Coh × Reward: activity with non-zero coefficients for the coherence × reward size interaction. Coh + Reward: activity with non-zero coefficients for coherence on trials with either choice and non-zero coefficients for either reward context or reward size.

(E, F) Activity of two example neurons. Shades: coherence levels. Colors: reward context (A) and reward size (B). Firing rates were computed using a 200 ms running window (50-ms steps). Only correct trials were included.



**Figure 3.3-figure supplement 1. Example neurons with reward context modulation.**

Rasters of the spiking activity of two example neurons. Cyan and red indicate transitions from Contra-large reward context to Ipsi-large reward context and from Contra-large reward context to Ipsi-large reward context, respectively. It is obvious that the neuron on the left is more active in Ipsi-large reward context; the neuron on the right is more active in Contra-large reward context. The change in neural activity coincided well with block transitions in both neurons.

***During motion-viewing: sensory and reward information are combined in some individual caudate neurons, but not in the format of decision variable.***

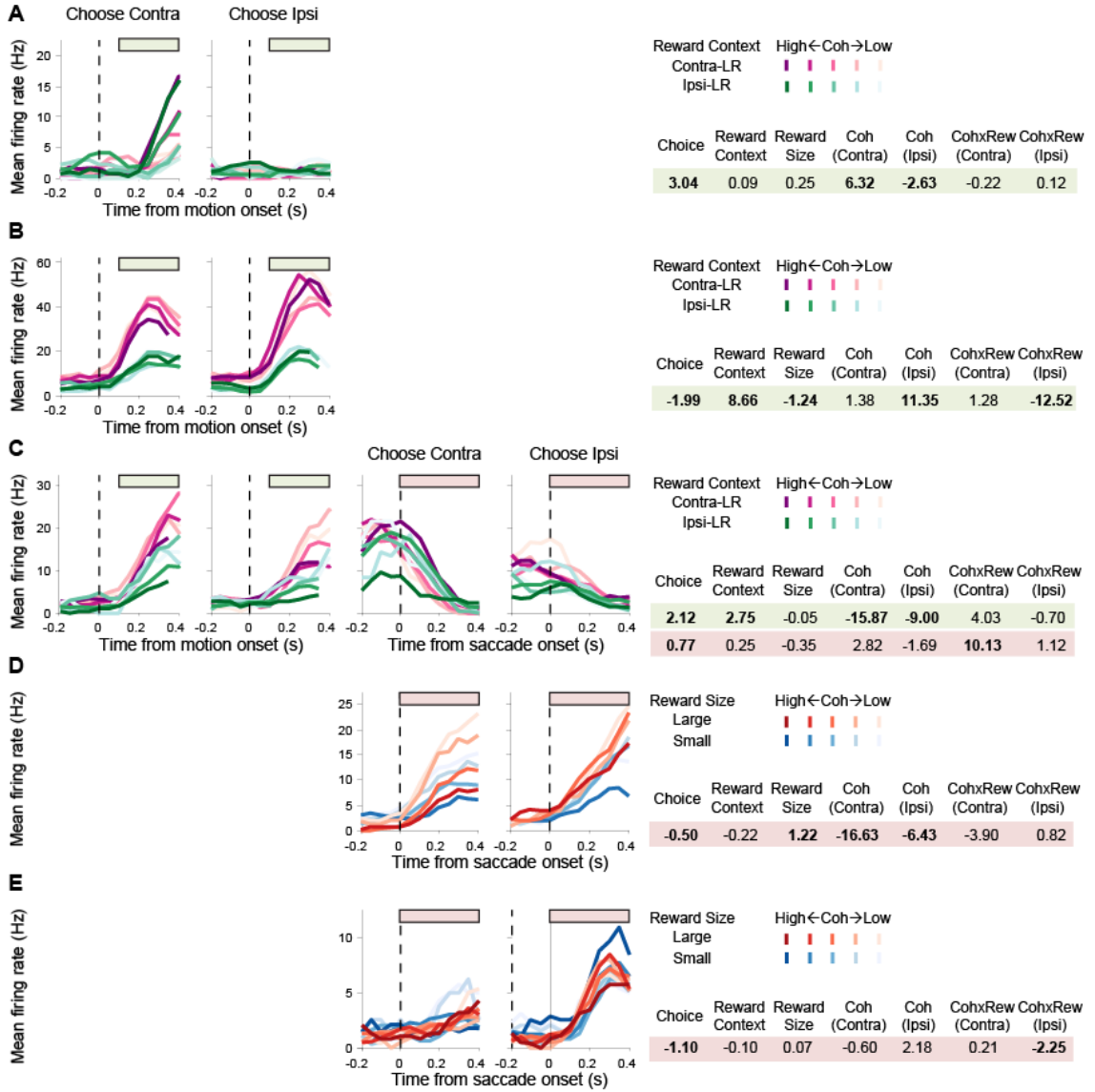
Previous studies have established the role of caudate nucleus in evidence accumulation, reward processing and decision formation (Nakamura and Hikosaka, 2006; Ding and Gold, 2010, 2012a; Yartsev et al., 2018). Are sensory information and reward-related information combined in individual caudate neurons? To answer this question, we searched for neurons with joint modulation by coherence and either reward context or reward size.

We found that many neurons with such joint modulation (Figure 3.3 E and F, Figure 3.3-figure supplement 2 A-C). For example, the activity of the neuron depicted in Figure 3.3E showed three types of modulation: 1) more activity during the blocks when the contralateral choice was paired with small reward and the ipsilateral choice was

paired with large reward (green > purple); 2) more activity for trials with stronger versus weaker motion evidence (dark shade > light shade; i.e., higher versus lower coherence levels, respectively), particularly for trials with contralateral choices; and 3) more activity for trials with contralateral versus ipsilateral choices, both during motion viewing and around saccade onset (Contra > Ipsi). This neuron's activity thus reflected a combination of reward context, motion strength, and eventual choice. The example neuron depicted in Figure 3.3F showed: 1) more activity on trials with higher coherence levels (dark shade > light shade); 2) a contralateral choice preference, both during motion viewing and around saccade onset (Contra > Ipsi); and 3) more activity when the choice was associated with large reward (red > blue). This neuron's activity thus reflected choice, the strength of motion stimulus leading to the choice and the reward size expected for the choice. In the caudate population, the presence of neurons with the joint modulation was also above chance level (Figure 3.3 D). This suggests that, sensory information and reward information are combined in some individual caudate neurons.

The combination of sensory and non-sensory information has been found in other brain areas, such as the lateral intraparietal area (LIP) (Rorie et al., 2010; Hanks et al., 2011). Hanks and colleagues found that LIP neural activity during decision-formation combined motion-direction evidence and the prior about how often the two alternatives could occur. Rorie and colleague found that LIP neural activity during decision-formation combined coherence and reward sizes of different options, similar to what we found in the caudate nucleus. In both studies, the LIP activity resembled the decision variable in the "accumulation to bound" framework. Is it possible that the caudate neural activity during decision formation also represents the decision variable?

## Single neuron examples



**Figure 3.3-figure supplement 2: Example neurons with different kinds of task-relevant modulations.**

Same format as Figure 3.3E and F. Bars above the curves indicate the epochs used for regression analysis results. Regression coefficients for each epoch are shown on the right. Bold: t-test,  $p < 0.05$ . Note that colors indicate reward contexts in A-C and reward size in D-E.

**(A)** A neuron with activity modulated by choice and coherence, but not reward-related quantities, during motion viewing.

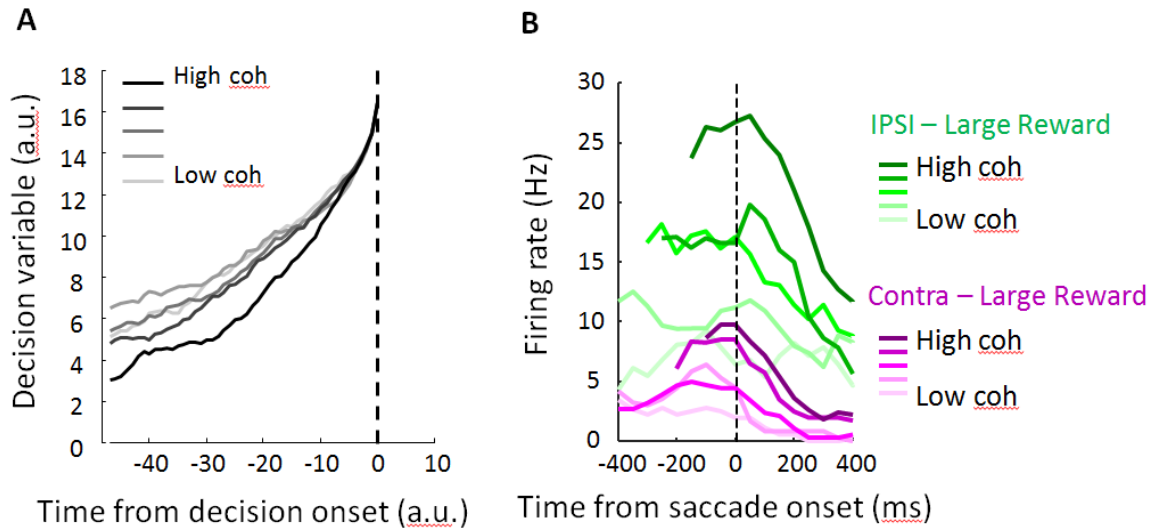
**(B)** A neuron with activity modulated by choice, reward context, expected reward size, coherence, and coherence-reward size interaction in trials with ipsilateral choices.

**(C)** A neuron with different modulation patterns for activity during and after motion viewing.

**(D)** A neuron with post-decision activity modulated by choice, expected reward size and coherence.

**(E)** A neuron with post-decision activity modulated by choice and coherence-reward size interaction in trials with ipsilateral choices.

A feature of decision variable in reaction time task is that, it converges to a common bound regardless of the sensory evidence strength (illustrated in Figure 3.3-figure supplement 3 A). Therefore, neural correlates of a decision variable should have neural activity reaching a common level when aligned to decision onset, such as reported from some neurons in LIP and in the frontal eye field (FEF) (Roitman and Shadlen, 2002; Ding and Gold, 2012b). However, we did not find this pattern in the 44 caudate neurons with joint modulation by coherence and reward based on the visual inspection of the neural activity (e.g. Figure 3.3 E and F, Figure 3.3-figure supplement C). The absence of the converge-to-bound pattern is consistent with a previous study using the same motion discrimination task but with equal reward (Ding and Gold, 2010). This suggests that sensory evidence and reward information are combined in the caudate nucleus but not in the format of decision variable. So what might be the function of such caudate neurons that combine sensory and reward information?



**Figure 3.3-figure supplement 3: Comparison between decision variable and caudate neural activity.**

**(A)** Simulated decision variable aligned to decision onset.

**(B)** Example neural activity aligned to saccade onset. Same neuron as in Figure 3.3 E.

The combination of sensory and reward information by the caudate neurons might play an important role in the reward-biased perceptual decision-making process. To explore this, in a subsequent study, we applied electric micro-stimulation to the caudate nucleus during the motion-viewing epoch (Doi et al., 2019). The micro-stimulation induced changes in the decision behavior that were different in the two reward contexts, suggesting that caudate neural activities during decision-making directly influences how sensory and reward information are combined. Further modeling analyses of the data showed that the micro-stimulation effect could be explained by coordinated changes in the drift-rate and the decision rule in the accumulation-to-bound framework. Taken together, these results suggest that the caudate nucleus does not represent the decision variable that directly links evidence accumulation to specific

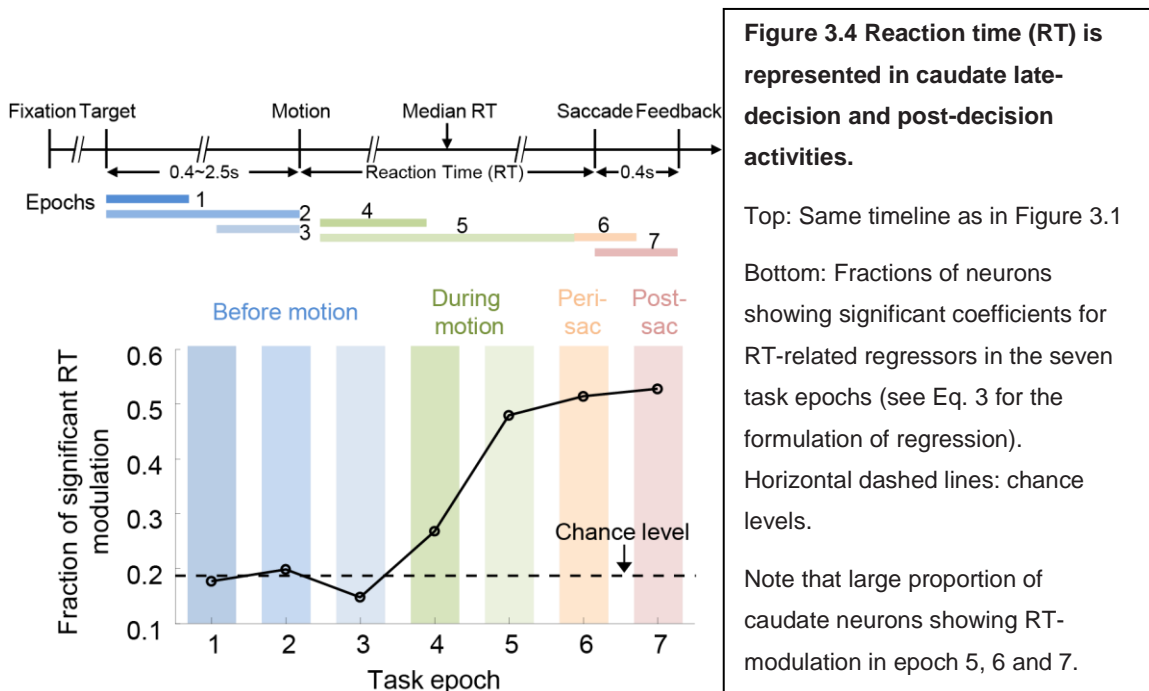
decisions, but rather modulates the kinds and magnitudes of biases and eventually shapes the decisions.

***Post-decision: sensory and reward information remains combined in caudate neurons, possibly for decision evaluation.***

During the decision formation epoch, we found that the caudate nucleus represented decision-related information, such as coherence, reward context, choice and reward size. In the post-decision period, this information was still represented and by even larger proportion of caudate neurons (Figure 3.3 A: peri-saccade (orange) and post-saccade (red) epochs). The proportion of neurons with joint modulation by reward and coherence was also slightly larger in the post-decision period than in the pre-decision period (57 and 62 for the peri- and post-saccade epochs, respectively, compared to 44 for the decision epoch). Of these neurons, 38% (in both peri- and post-saccade epochs) preferred high coherence for one choice and low coherence for the other (e.g. Figure 3.3-figure supplement 1 C), providing a memory of the amount of evidence supporting the specific choice. The other 62% preferred high or low coherence for both choices (e.g. Figure 3.3 F and Figure 3.3-figure supplement 1 D). These neurons could reflect the difficulty of the decision, because task difficulty is independent of the specific choice being made. This kind of choice-independent coherence-modulation is consistent with the finding in a previous study using the equal-reward version of the same task. In addition, we found that, during this time period, a significant population of caudate neurons also represented the reaction time (RT) of the decision (Figure 3.4). Because reaction time is modulated by both coherence and reward size



(Chapter 2), these results suggest that sensory evidence and reward information were combined in the post-decision caudate neural activity as well.



Unlike the neural activity during decision-making, post-decision activity could no longer influence the current decision. However, our results, as well those from several other studies, have found that caudate nucleus continues representing task- and behavior-related variables after decisions, such as task difficulty and choice value (Lau and Glimcher, 2008; Ding and Gold, 2010; Yanike and Ferrera, 2014). It is possible that information represented during this period could be used for behavioral monitoring and evaluation.

This hypothesis is plausible, because the choice-independent coherence modulation, described in the paragraph above, could provide information about task difficulty, which is important for evaluation. In RT task, decision accuracy has been found to decrease with decision time, because longer decision times tend to associated with

more difficult trials(Hanks et al., 2011). Therefore, the RT-modulated neurons could provide information about decision accuracy. In addition, reward expectation could be computed by combining decision accuracy with the reward size of the choice (which can be provided by the reward size-modulated neurons). In sum, sensory evidence and reward information were combined in individual caudate neurons, which could be used for decision evaluation.

## **Discussion**

We recorded neural activity in the caudate nucleus of two monkeys making reward-biased perceptual decisions. We found that multiple task- and decision-related features were represented before, during and after the decision. The emergence of these features was consistent with the timeline of the task. For example, reward context was always available throughout the trial, and we found reward context-representation before, during and after the decision. Motion coherence was only represented after motion stimulus onset. Choice and its associated reward size representations emerged as sensory evidence was accumulated and combined with reward information to form a decision.

We also found single neurons that jointly represented sensory and reward information both during and after decision, which supports the hypothesis that caudate combines multiple sources of information in support of a choice.

### ***Information representation before stimulus onset***

We found that before the presentation of sensory information, caudate nucleus represented the reward context information, a phenomenon that has also been observed

in other studies with a similar block-design for reward context manipulation (Lauwereyns et al., 2002; Ding and Hikosaka, 2006). This reward context information could be used to establish a bias in the baseline of evidence accumulation, similar to changing the starting value or decision threshold in the accumulation-to-bound model. In the context of the computational model, similar bias (prior) has been shown to be induced by manipulating stimulus probability (Hanks et al., 2011; Mulder et al., 2012), although those studies have not recorded from caudate neurons during such manipulation. It would be interesting to know whether the caudate nucleus only represents reward-related bias or could also represent sensory prior.

### ***Information representation during decision***

During decision, we found that sensory evidence and reward information are combined in the caudate nucleus, which could be used for decision formation. Subsequent micro-stimulation experiments have confirmed the direct involvement of caudate nucleus in combining sensory and reward information when making complex decisions (Doi et al., 2019). However, as shown in the results section, the activity of caudate neurons that jointly represented sensory evidence and reward information did not resemble a decision variable that directly links evidence accumulation with specific decisions. Meanwhile, other brain areas in the parietal and prefrontal cortex have been found to compute decision variables that represent the combination of sensory and non-sensory information. In monkeys, decision variable-like signals that combines sensory evidence with reward-bias or prior, have been found in area LIP (Rorie et al., 2010; Hanks et al., 2011). In human, Model-based fMRI studies found decision variable-like BOLD signals in areas like the inferior parietal lobule, superior parietal lobule, lateral

frontopolar and orbitofrontal cortices, although without the temporal precision of single-unit recording (Summerfield and Koechlin, 2010).

From these findings, it appears that the decision-making process is complex and involves multiple brain areas. Our results suggest that the role of the caudate nucleus might be to encode non-decision variable information. Such signals could then modulate the balance between sensory evidence and reward bias and influence the evolution of decision variables in other parts of the brain.

### ***Information representation after decision***

We found that after a decision was made, sensory and reward information were still represented in the caudate nucleus. Such information can no longer influence the current decision, but they might be used for decision performance evaluation. Examples of such evaluation include: (a) computation of task difficulty from neurons with coherence-modulation; (b) computation of decision accuracy from neurons with RT-modulation and (c) computation of reward expectation from neurons with both coherence- and reward-modulation.

In other brain areas, some post-action neural activities have been previously reported to represent performance monitoring quantities. For example, in the frontal eye field, some neurons' post-decision activity was found to correlate with the correctness and difficulty of current trials in a speed categorization task (Teichert et al., 2014). In a motor learning task, the post-action neural activity in monkey area LIP was found to encode the error of the motor execution (Zhou et al., 2016). Post-decision neural activity in monkey ventral lateral prefrontal cortex (vLPFC) was found to encode information related to the current decision as well as choice bias in the next decision in an auditory

decision task(Tsunada et al., 2019). Peri-action activity in the prelimbic region of rat medial prefrontal cortex (PL) correlated with expected value(Lak et al., 2019). In both monkey vIPFC and rat PL, perturbing the neural activity via microstimulation and optogenetic silencing, respectively, did not influence the current decision, but influenced behavior in subsequent trials, suggesting that these monitoring signal might guide behavioral adjustments in the future. It is possible that, like these brain areas, the caudate post-decision activity might be used for evaluating current performance and adjusting future actions.

In the next chapter, I will examine this hypothesis in detail by focusing on whether and how caudate post-decision activity might represent two evaluative quantities—confidence and reward expectation.

## **Materials and methods**

### ***Subjects***

Two of the three monkeys in Chapter 2 (monkey C and monkey F) were used for the experiments in this Chapter.

### ***Behavioral task***

Same as described in Chapter 2.

### ***Data acquisition***

Eye position was monitored using a video-based system (ASL) sampled at 240 Hz. Single-unit recordings focused on putative project neurons (Ding and Gold, 2010). We searched for task-relevant neurons while the monkeys performed the equal-reward motion discrimination task with horizontal dots motions and determined the presence of task-related modulation of neural activity by visual and audio inspection of ~10–20 trials.

For analyses of neural response properties in recording sessions, only well-isolated single units were included. Neural signals were amplified, filtered and stored using a MAP acquisition system (Plexon, Inc.), along with time-stamped event codes, analog eye position signals and trial parameter values. Single unit activity was identified by offline spike sorting (Offline Sorter, Plexon, Inc.).

### ***Neural data analysis***

For each single unit dataset, we computed the average firing rates in seven task epochs (Figure 3.1A): three epochs before motion stimulus onset (400 ms window beginning at target onset, variable window from target onset to dots onset, and 400 ms window ending at motion onset), two epochs during motion viewing (a fixed window from 100 ms after motion onset to 100 ms before median RT and a variable window from 100 ms after motion onset to 100 ms before saccade onset), a peri-saccade 300 ms window beginning at 100 ms before saccade onset, and a post-saccade 400 ms window beginning at saccade onset (before feedback and reward delivery).

### ***Identify reward context modulation before motion stimulus onset (Figure 3.2):***

For each unit, the following multiple linear regression was performed on the average firing rates in epoch 2 in all trials.

$$FR = \beta_0 + \beta_{\text{Choice}} \times I_{\text{Choice}} + \beta_{\text{RewCont}} \times I_{\text{RewCont}} + \beta_{\text{RewSize}} \times I_{\text{RewSize}} \quad (\text{Eq.1})$$

$$\text{where } I_{\text{Choice}} = \begin{cases} 1 & \text{for contralateral choice} \\ -1 & \text{for ipsilateral choice} \end{cases}$$

$$I_{\text{RewCont}} = \begin{cases} 1 & \text{for contralateral – large reward blocks} \\ -1 & \text{for ipsilateral – large reward blocks} \end{cases}$$

$$I_{\text{RewSize}} = \begin{cases} 1 & \text{if a large reward is expected for the choice} \\ -1 & \text{if a small reward is expected for the choice} \end{cases}$$

Significance of non-zero coefficients was assessed using  $t$ -test (criterion:  $p=0.05$ ).

Identify directly measured decision-related modulations (Figure 3.3 A):

For each unit, the following multiple linear regression was performed on the average firing rates in correct trials for each task epoch separately.

$$\begin{aligned} \text{FR} = & \beta_0 + \beta_{\text{Choice}} \times I_{\text{Choice}} + \beta_{\text{RewCont}} \times I_{\text{RewCont}} + \beta_{\text{RewSize}} \times I_{\text{RewSize}} \\ & + \beta_{\text{Coh-Contr}} \times I_{\text{Coh-Contr}} + \beta_{\text{Coh-Ipsi}} \times I_{\text{Coh-Ipsi}} \\ & + \beta_{\text{RewCoh-Contr}} \times I_{\text{Coh-Contr}} \times I_{\text{RewSize}} + \beta_{\text{RewCoh-Ipsi}} \times I_{\text{Coh-Ipsi}} \times I_{\text{RewSize}}, \end{aligned}$$

(Eq. 2)

$$\text{where } I_{\text{Choice}} = \begin{cases} 1 & \text{for contralateral choice} \\ -1 & \text{for ipsilateral choice} \end{cases}$$

$$I_{\text{RewCont}} = \begin{cases} 1 & \text{for contralateral – large reward blocks} \\ -1 & \text{for ipsilateral – large reward blocks} \end{cases}$$

$$I_{\text{RewSize}} = \begin{cases} 1 & \text{if a large reward is expected for the choice} \\ -1 & \text{if a small reward is expected for the choice} \end{cases}$$

$$I_{\text{Coh-Contr}} = \begin{cases} \text{absolute coherence for contralateral choice (centered at mean value)} \\ 0 & \text{for ipsilateral choice} \end{cases},$$

$$\text{and } I_{\text{Coh-Ipsi}} = \begin{cases} 0 & \text{for contralateral choice} \\ \text{absolute coherence for ipsilateral choice (centered at mean value)} \end{cases}$$

Significance of non-zero coefficients was assessed using  $t$ -test (criterion:  $p=0.05$ ).

Identify RT-related modulations (Figure 3.4):

For each unit, the following multiple linear regression was performed on the average firing rates in all trials for each task epoch separately.

$$\begin{aligned}
 FR = & \beta_0 + \beta_{\text{Choice}} \times I_{\text{Choice}} + \beta_{\text{RewCont}} \times I_{\text{RewCont}} + \beta_{\text{RewSize}} \times I_{\text{RewSize}} \\
 & + \beta_{\text{RT-Contralateral}} \times I_{\text{RT-Contralateral}} + \beta_{\text{RT-Ipsilateral}} \times I_{\text{RT-Ipsilateral}} \\
 & + \beta_{\text{RewRT-Contralateral}} \times I_{\text{RT-Contralateral}} \times I_{\text{RewSize}} + \beta_{\text{RewRT-Ipsilateral}} \times I_{\text{RT-Ipsilateral}} \times I_{\text{RewSize}},
 \end{aligned}
 \tag{Eq. 3}$$

where  $I_{\text{Choice}} = \begin{cases} 1 & \text{for contralateral choice} \\ -1 & \text{for ipsilateral choice} \end{cases}$ ,

$I_{\text{RewCont}} = \begin{cases} 1 & \text{for contralateral – large reward blocks} \\ -1 & \text{for ipsilateral – large reward blocks} \end{cases}$ ,

$I_{\text{RewSize}} = \begin{cases} 1 & \text{if a large reward is expected for the choice} \\ -1 & \text{if a small reward is expected for the choice} \end{cases}$ ,

$I_{\text{Coh-Contralateral}} = \begin{cases} \text{RT for contralateral choice (centered at mean value)} \\ 0 & \text{for ipsilateral choice} \end{cases}$ ,

and  $I_{\text{Coh-Ipsilateral}} = \begin{cases} 0 & \text{for contralateral choice} \\ \text{RT for ipsilateral choice (centered at mean value)} \end{cases}$ .

Significance of non-zero coefficients was assessed using  $t$ -test (criterion:  $p=0.05$ ).

RT-modulated neurons were identified as neurons with significant  $\beta_{\text{RT-Contralateral}}$ ,  $\beta_{\text{RT-Ipsilateral}}$ ,

$\beta_{\text{RewRT-Contralateral}}$  or  $\beta_{\text{RewRT-Ipsilateral}}$ .



### ***Behavioral analysis***

A logistic function was fitted to the choice data for all trials:

$$P_{\text{contra choice}} = \frac{1}{1 + e^{-\text{Slope} \times (\text{Coh} + \text{Bias})}} \quad , \quad (\text{Eq. 4})$$

Where Coh is the signed motion coherence,

$$\text{Slope} = \text{slope}_0 + \text{slope}_{\text{rew}} \times \text{RewCont} \quad ,$$

$$\text{Bias} = \text{bias}_0 + \text{bias}_{\text{rew}} \times \text{RewCont} \quad ,$$

$$\text{RewCont} = \begin{cases} 1 & \text{for contralateral – large reward blocks} \\ -1 & \text{for ipsilateral – large reward blocks} \end{cases} \quad ,$$

## Reference

- Bogacz R, Gurney K (2007) The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput* 19:442–477.
- Cai X, Kim S, Lee D (2011) Heterogeneous Coding of Temporally Discounted Values in the Dorsal and Ventral Striatum during Intertemporal Choice. *Neuron* 69:170-182.
- Cavanagh JF, Wiecki T V., Cohen MX, Figueroa CM, Samanta J, Sherman SJ, Frank MJ (2011) Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nat Neurosci* 14:1462–1467.
- Diederich A, Busemeyer JR (2006) Modeling the effects of payoff on response bias in a perceptual discrimination task : two-stage-processing hypothesis. *Percept Psychophysics* 68:194–207.
- Ding L, Gold JI (2010) Caudate encodes multiple computations for perceptual decisions. *J Neurosci* 30:15747–15759.
- Ding L, Gold JI (2012a) Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. *Neuron* 75:865–874.
- Ding L, Gold JI (2012b) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cereb Cortex* 22:1052–1067.
- Ding L, Gold JI (2013) The basal ganglia's contributions to perceptual decision making. *Neuron* 79:640–649.
- Ding L, Hikosaka O (2006) Comparison of reward modulation in the frontal eye field and caudate of the macaque. *J Neurosci* 26:6695–6703.
- Doi T, Fan Y, Gold JI, Ding L (2019) The caudate nucleus controls coordinated patterns of adaptive, context-dependent adjustments to complex decisions. *bioRxiv*:568733.

- Fan Y, Gold JI, Ding L (2018) Ongoing, rational calibration of reward-driven perceptual biases. *Elife* 7:e36018.
- Feng S, Holmes P, Rorie A, Newsome WT (2009) Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Comput Biol* 5:e1000284.
- Gao J, Tortell R, McClelland JL (2011) Dynamic integration of reward and stimulus information in perceptual decision-making. *PLoS One* 6:e16749.
- Hanks TD, Mazurek ME, Kiani R, Hopp E, Shadlen MN (2011) Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J Neurosci* 31:6339–6352.
- Hikosaka O, Kim HF, Yasuda M, Yamamoto S (2014) Basal Ganglia Circuits for Reward Value–Guided Behavior. *Annu Rev Neurosci* 37:289–306.
- Hikosaka O, Sakamoto M, Usui S (1989) Functional properties of monkey caudate neurons. I. Activities related to saccadic eye movements. *J Neurophysiol* 61:780–798.
- Kable JW, Glimcher PW (2009) The neurobiology of decision: consensus and controversy. *Neuron* 63:733–745.
- Kiani R, Hanks TD, Shadlen MN (2008) Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci* 28:3017–3029.
- Kim HF, Hikosaka O (2013) Distinct Basal Ganglia Circuits Controlling Behaviors Guided by Flexible and Stable Values. *Neuron* 79:1001–1010.
- Kimchi EY, Laubach M (2009) The Dorsomedial Striatum Reflects Response Bias during Learning. *J Neurosci* 29:14891–14902.
- Lak A, Okun M, Moss M, Gurnani H, Farrell K, Wells MJ, Reddy CB, Kepecs A, Harris

- KD, Carandini M (2019) Neural basis of learning guided by sensory confidence and reward value. *bioRxiv*:411413.
- Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58:451–463.
- Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002) A neural correlate of response bias in monkey caudate nucleus. *Nature* 418:413–417.
- Leite FP, Ratcliff R (2011) What cognitive processes drive response biases? A diffusion model analysis. *Judgm Decis Mak* 6:651–687.
- Liston DB, Stone LS (2008) Effects of prior information and reward on oculomotor and perceptual choices. *J Neurosci* 28:13866–13875.
- Maddox WT, Bohil CJ (1998) Base-rate and payoff effects in multidimensional perceptual categorization. *J Exp Psychol Learn Mem Cogn* 24:1459–1482.
- Mulder MJ, Wagenmakers E-J, Ratcliff R, Boekel W, Forstmann BU (2012) Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J Neurosci* 32:2335–2343.
- Nakamura K, Hikosaka O (2006) Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *J Neurosci* 26:5360–5369.
- Rao RPN (2010) Decision making under uncertainty: a neural model based on partially observable markov decision processes. *Front Comput Neurosci* 4:146.
- Ratcliff R, Frank MJ (2012) Reinforcement-Based Decision Making in Corticostriatal Circuits: Mutual Constraints by Neurocomputational and Diffusion Models. *Neural Comput* 24:1186–1229.
- Redgrave P, Prescott TJ, Gurney K (1999) The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience* 89:1009–1023

- Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22:9475–9489
- Rorie AE, Gao J, McClelland JL, Newsome WT (2010) Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS One* 5:e9308
- Samejima K, Doya K (2007) Multiple representations of belief states and action values in corticobasal ganglia loops. *Ann N Y Acad Sci* 1104:213–228
- Santacruz SR, Rich EL, Wallis JD, Carmena JM (2017) Caudate Microstimulation Increases Value of Specific Choices. *Curr Biol* 27:3375-3383.e3.
- Seo M, Lee E, Averbeck BB (2012) Action Selection and Action Value in Frontal-Striatal Circuits. *Neuron* 74:947–960.
- Silkis I (2001) The cortico-basal ganglia-thalamocortical circuit with synaptic plasticity. II. Mechanism of synergistic modulation of thalamic activity via the direct and indirect pathways through the basal ganglia. *Biosystems* 59:7–14
- Summerfield C, Koechlin E (2010) Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. *Front Hum Neurosci* 4:208.
- Summerfield C, Tsetsos K (2012) Building Bridges between Perceptual and Economic Decision-Making: Neural and Computational Mechanisms. *Front Neurosci* 6:70
- Tachibana Y, Hikosaka O (2012) The Primate Ventral Pallidum Encodes Expected Reward Value and Regulates Motor Action. *Neuron* 76: 826-837.
- Tai LH, Lee AM, Benavidez N, Bonci A, Wilbrecht L (2012) Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat Neurosci* 15:pages1281–1289.

- Teichert T, Yu D, Ferrera VP (2014) Performance monitoring in monkey frontal eye field. *J Neurosci* 34:1657–1671.
- Tsunada J, Cohen Y, Gold JI (2019) Post-decision processing in primate prefrontal cortex influences subsequent choices on an auditory decision-making task. *Elife* 8:1–21.
- Voss A, Rothermund K, Voss J (2004) Interpreting the parameters of the diffusion model: an empirical validation. *Mem Cognit* 32:1206–1220
- Waiblinger C, Wu CM, Bolus MF, Borden PY, Stanley GB (2019) Stimulus Context and Reward Contingency Induce Behavioral Adaptation in a Rodent Tactile Detection Task. *J Neurosci* 39:1088–1099
- Wang L, Rangarajan K V., Gerfen CR, Krauzlis RJ (2018) Activation of Striatal Neurons Causes a Perceptual Decision Bias during Visual Change Detection in Mice. *Neuron* 97:1369-1381.e5.
- Whiteley L, Sahani M (2008) Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *J Vis* 8:2.1-15
- Yanike M, Ferrera VP (2014) Interpretive monitoring in the caudate nucleus. *Elife* 3:1–16
- Yartsev MM, Hanks TD, Yoon AM, Brody CD (2018) Causal contribution and dynamical encoding in the striatum during evidence accumulation. *Elife* 7:1–24.
- Zhou Y, Liu Y, Lu H, Wu S, Zhang M (2016) Neuronal representation of saccadic error in macaque posterior parietal cortex (PPC). *Elife* 5:1–17.

## CHAPTER 4: CONFIDENCE AND REWARD EXPECTATION ARE REPRESENTED IN CAUDATE POST-DECISION ACTIVITY

Yunshu Fan, Takahiro Doi, Joshua I. Gold, Long Ding

### Introduction

In chapter 3, I discussed the possibility that the caudate post-decision activity could carry information for decision monitoring and evaluation. In this chapter, I will focus on examining whether the caudate neurons could represent two specific evaluative quantities – confidence and reward expectation.

Confidence is the subjective belief, prior to feedback, that a decision is correct (Kiani et al., 2014). It is particularly relevant in the context of making a decision based on unreliable or noisy evidence, and it could influence how to act subsequently upon the current decision. For example, I see a dark patch on the ground in front of me. Knowing that my eyesight is very good, I decide quite confidently that it is some darker-colored soil, not a puddle of water, so I know stepping on it would be fine. If my eyesight is pretty bad, after staring at it for a while, I might still reach the same conclusion, but my confidence of that conclusion would be quite low, and I might recommend people to jump over it, in case it is a puddle of water. Studies on confidence in monkeys making categorical judgement on noisy sensory stimulus usually set up as such: in some trials, in addition to the two perceptual categories, monkeys were given the chance to choose a third safe option that guarantees a reward smaller than the amount they would get if they pick the correct choice. They found that those monkeys were more likely to choose

the option with the guaranteed smaller reward when the monkeys were less confident about the stimuli (Kiani and Shadlen, 2009; Fetsch et al., 2014). Other studies in both monkeys and rats found that they were more likely to abort an uncertain decision in order to reinitiate a new trial (Kepecs et al., 2008; Yanike and Ferrera, 2014). Post-decision confidence could provide information about how the subject should adjust subsequent behavior. For example, uncertainty could modulate the learning rate used for belief updating in changing environments modulate learning rate (Yu and Dayan, 2005; Nassar et al., 2012).

Reward expectation is the product of the probability of obtaining a reward and the magnitude of the reward. According to the expected utility theory, in value-based decision-making paradigm, optimal decision should favor the option with the higher reward expectation (Rangel et al., 2008). This theory has been verified in animal matching behavior tasks (Lau and Glimcher, 2008). Post-decision reward expectation could be used for computing the “reward prediction error”, a key quantity in the reinforcement learning framework (Sutton and Barto, 1998; Samejima et al., 2005; Daw and Doya, 2006; Schultz, 2015). The reward prediction error, hence reward expectation, is also useful for detecting environment change in order to adapt the behavior and strategy accordingly (Behrens et al., 2007; Nassar et al., 2010; Mathys et al., 2011; Meder et al., 2017).

Confidence and reward expectation are closely linked, because, from the decision maker’s point of view, the probability of getting the reward is essentially the estimation of the probability of a decision being made is correct, in other words, confidence. Therefore, reward expectation becomes a scaled version of confidence, with reward magnitude being the scalar. When there is no internal bias, confidence of



choosing each option should be the same. When the two options are associated with the same magnitude of reward, confidence and reward expectation are perfectly correlated. This perfect correlation between confidence and reward expectation poses challenge to distinguish their individual behavior effects and neural correlates. Therefore, confidence and reward expectation are usually not distinguished in the same study.

With our reward-biased visual motion discrimination task (the same task as in Chapter 2 and 3), confidence is no longer perfectly correlated with reward expectation, thus allowing us to differentiate their individual effects and neural representations. In this chapter, I will first present how we compute confidence and reward expectation and key features of these quantities under equal-reward, no-bias condition. Then I will show how reward asymmetry and reward bias enable us to differentiate confidence from reward expectation. Finally, I will examine the behavioral effects of confidence and reward expectation in our monkeys and the neural representations of these quantities in the caudate nucleus.

## **Results**

### ***Reward asymmetry-induced bias can help distinguish between confidence and reward expectation.***

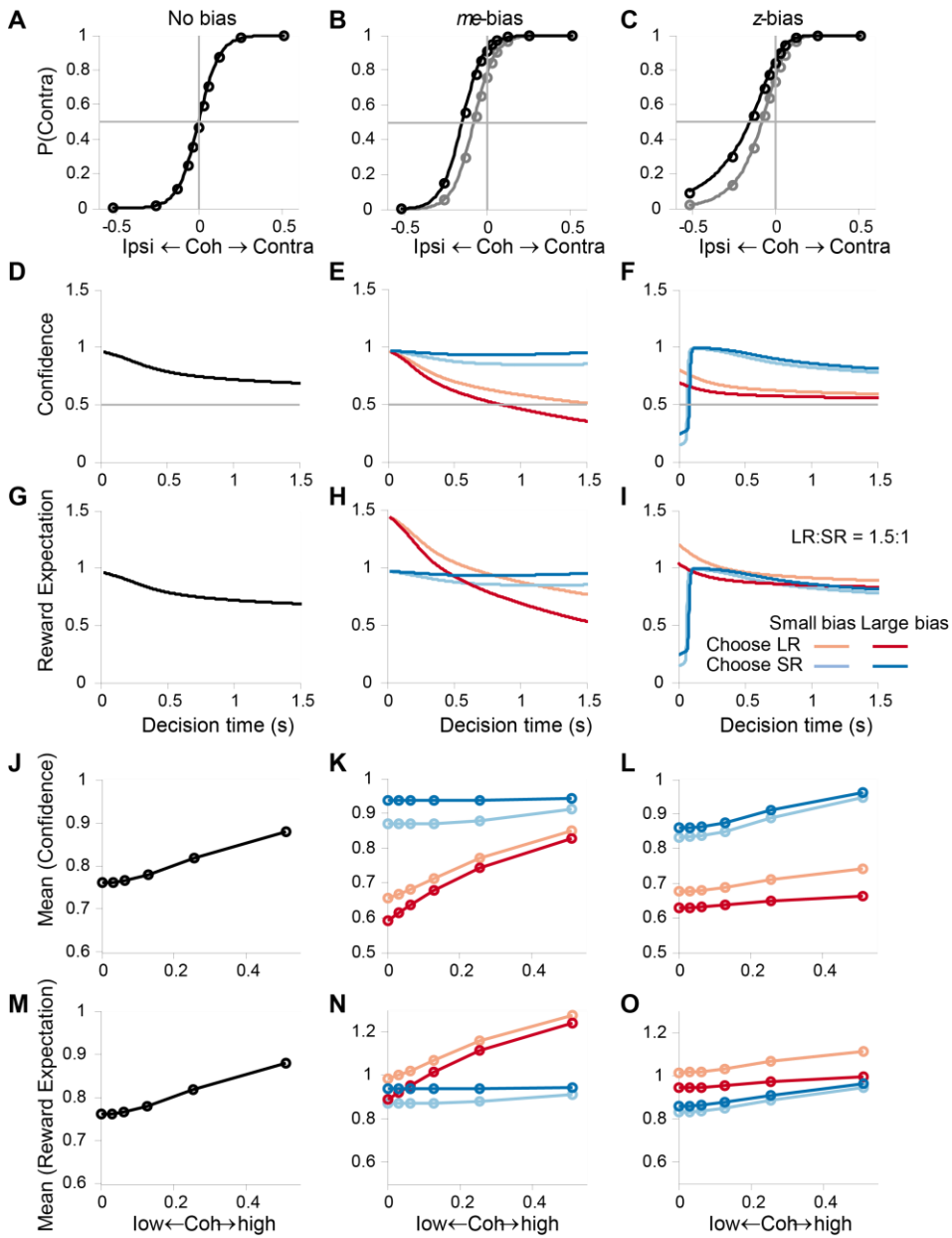
Because we did not have direct measurements of the monkeys' confidence levels, we first computed confidence and reward expectation from the monkeys' behavioral performance, based on the following assumptions: (1) the monkeys' decision processes were approximated by the drift-diffusion model as described in Chapter 2; (2) the monkeys did not have direct access to the motion coherence in each trial, because all coherence levels were randomly interleaved; and 3) their confidence depended on

the choice they made and the time it took to make that decision, marginalized over all possible coherence levels. Reward expectation was computed as a product of confidence and reward magnitude of the chosen option (for equal-reward condition, the reward magnitudes were set to 1; for asymmetric-reward condition, the reward magnitudes were normalized by the magnitude of the small reward).

We then verify that the confidence we computed follows the same pattern as the confidence measured in previous studies. Under the no-bias condition, the confidence we computed decreases as a function of decision time and increases as a function of motion coherence (Figure 4.1 D and J). These patterns are consistent with previous results based on confidence measured directly in human subjects performing an equal-reward version of our decision task (Kiani et al., 2014: Figure 2) or inferred from behavior performance (Kepecs et al., 2008; Kiani and Shadlen, 2009; Kiani et al., 2014; Lak et al., 2019). As expected under equal-reward condition, the reward expectation follows the same pattern as confidence (Figure 4.1 G and M), making the two quantities indistinguishable.

In contrast, with reward asymmetry-induced biases, confidence and reward expectation are no longer perfectly correlated (Figure 4.1 middle and right columns). Confidence for small-reward choices is overall higher than that for large-reward choices (Figure 4.1E, F, K and L: blue curves are above red curves), because a small-reward choice requires more sensory evidence to support it, therefore is more likely to be correct. This difference in confidence increases with bias (compare the distance between red and blue curves for dark and light shades). Reward expectation, on the other hand, can be lower for small-reward choices with high confidence than for large-reward choices with low confidence, depending on the reward ratio (Figure 4.1 H, I, N

and O: the blue curves are under the red curves). Given the same reward ratio, the difference in reward expectation may decrease with bias. These results suggest that, the presence of reward-bias in the decision-making process and the difference in reward magnitude of the two choices reduce the correlation between confidence and reward expectation, making them partially distinguishable in the same experiment.



**Figure 4.1. Confidence and reward expectation depended on decision time, motion coherence, and reward asymmetry-induced biases.**

**(A- C)** Psychometric functions of DDM- simulated decision behaviors in three scenarios: no-bias, equal reward (A), *me*-bias to contralateral (large reward) option (B), *z*-bias to contralateral (large reward) option (C). Circles indicate the coherence levels that were interleaved in the simulated trials. Black/gray color in B and C: larger/smaller reward biases. DDM parameters used are:  $a=2$ ;  $k=8$ ;  $t_{nd} = 0$ ; fixed bound;  $me = 0.08$  and  $0.15$  for smaller and larger *me*-bias, respectively;  $z = 0.76$  and  $0.86$  for smaller and larger *z*-bias, respectively. *me* and *z* were chosen to generate similar amount of choice-bias (horizontal shift of the psychometric function at chance level).

**(D-F)** Confidence as a function of decision time, computed from the simulated behaviors in A-C, respectively. Red/Blue: trials choosing the large/small reward options. Darker/lighter shades: behavior with larger/smaller reward bias (corresponds to the black/gray psychometric functions in B and C).

**(G-I)** Reward expectation as a function of decision time. Same format as D-F. LR: large reward; SR: small reward. Reward magnitude = 1.5 and 1 for large and small reward options, respectively.

**(J-O)** Confidence (in J-L) and reward expectation (in M-O) as a function of motion coherence, computed from the simulated behaviors in A-C, respectively. Same format as D-F.

Because reward-bias in the DDM could be generated by two different mechanisms: biasing the drift-rate (*me*-bias) and biasing the decision-rule (*z*-bias), and because the monkeys used both kinds of biasing mechanisms (as shown in Chapter 2), we examined how *me*-bias and *z*-bias influence confidence and reward expectation respectively. For confidence, *me*-bias tends to magnify the difference between small- and large-reward options more at longer decision time and for lower coherences (Figure 4.1 E and K). In contrast, *z*-bias tends to increase the confidence difference between small- and large-reward options more at shorter decision time and for higher coherences (Figure 4.1 F and L). For reward expectation, the patterns are almost the opposite: *me*-bias tends to magnify the difference between small- and large-reward options more at shorter decision time and for higher coherences (Figure 4.1 H and N), whereas *z*-bias tends to increase the confidence difference between small- and large-reward options

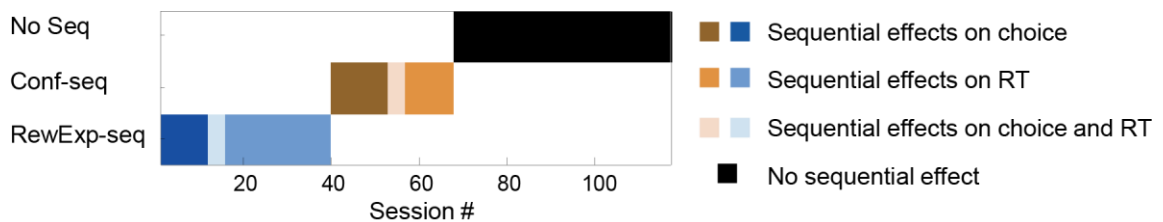
more at longer decision time and for lower coherences (Figure 4.1 I and O). These different patterns further suggested that confidence and reward expectation are two distinguishable quantities in our experiment.

***Confidence and reward expectation influenced monkeys' subsequent decision behavior in some sessions***

Confidence and reward expectation can be used in conjunction with feedback/reward outcome to evaluate how well the decision was made and whether adjustments in subsequent decisions are necessary. Inspired by a recent study demonstrating confidence-dependent post-error adjustments in well-trained monkeys (Kiani's post-error paper), we examined whether the monkeys on our task used confidence or reward expectation to adjust their subsequent decisions. In other words, we assessed the degree to which each evaluative quantity computed from the previous trial affected the monkeys' choice and reaction time for the current trial. Due to the small numbers of error trials, we focused our analysis of these sequential effects on only decisions following correct trials.

We used model fitting with logistic functions to measure two potential sequential effects on the monkeys' choice behavior: (1) increase or decrease the tendency to choose the large-reward options ("reward bias"), reflected as opposite-direction shifts in the psychometric functions of the two reward contexts (Figure 4.2 B, right panel), and (2) increase or decrease the tendency to choose the contralateral option, regardless of its reward size ("choice bias"), reflected as same-direction shifts in the psychometric functions of the two reward contexts (Figure 4.2 C, right panel). We used model fitting with linear functions to measure two potential sequential effects on the monkeys' RT

behavior: (1) speed up or slow down the reaction time overall for both large- and small-reward choices (“baseline RT”), reflected as same-direction shifts (Figure 4.2 D, right panel) and (2) increase or decrease the difference in RT between large- and small-reward choices (“reward bias in RT”), reflected as opposite-direction shifts of the RT function for large- and small-reward choices (Figure 4.2 E, right panel). For each of the two evaluative quantities, we compared the goodness-of-fits of four models in order to identify the specific kinds of sequential effect: 1) “full model”: including effects on choice bias, reward bias, baseline RT, and reward bias in RT; 2) “choice-only”: including only effects on choice and reward biases; 3) “RT-only”: including only effects on baseline RT and reward bias in RT; and 4) “No-seq”: no sequential effects. We then compared the best-fitting models with confidence-dependent sequential effects and reward expectation-dependent sequential effects to assess whether the sequential effects in the monkeys’ behaviors were more likely to be confidence- or reward expectation-related.



**Figure 4.2 Confidence- and reward expectation-related sequential influence on monkeys’ choice and RT.**

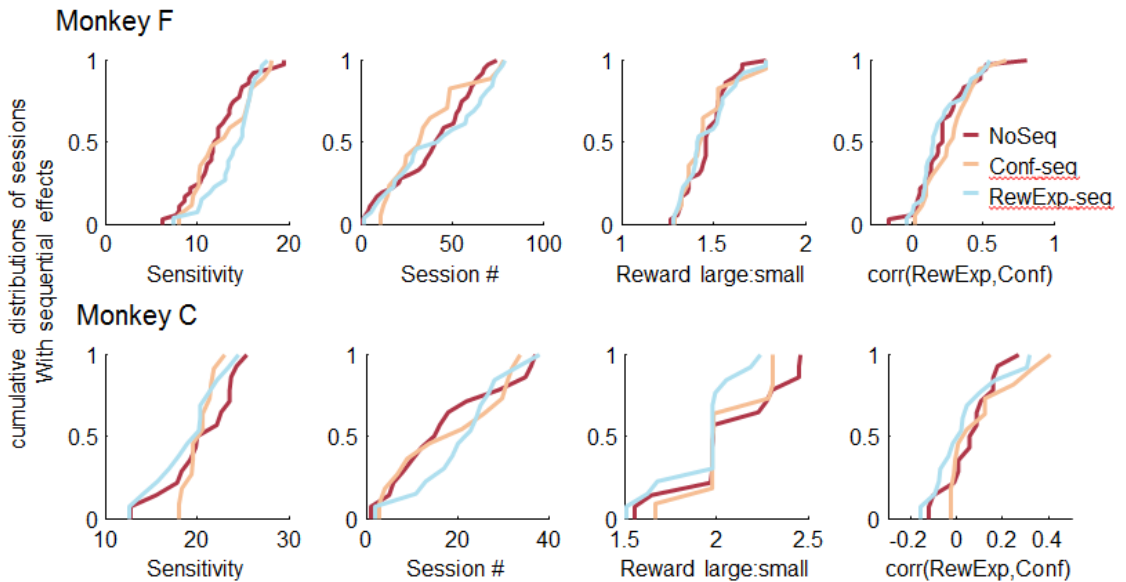
Sessions with no sequential effect (top row, black), confidence-related sequential effects (middle row, “Conf-seq”) and reward expectation-related sequential effects (bottom row, “RewExp-seq”). Shades: sessions with sequential effects on choice alone/ choice and RT/ RT alone.

Of the 117 sessions, we found confidence- and reward expectation-related sequential effects in 28 and 39 sessions, respectively. For each session, the sequential effect could be on choice alone, RT alone or both (different shades in the middle and

bottom rows in Figure 4.2 A). This suggested that both confidence and reward expectation influenced monkeys' subsequent decisions in subsets and separate sessions.

We then examined if any task or behavior parameter could predict whether the monkey had confidence-related, reward expectation-related, or no sequential effects. For example, we hypothesized that the monkeys tended to have no sequential effects when their perceptual sensitivity is high. To examine this hypothesis, we examine whether the cumulative distributions of sessions with no sequential effects (Figure 4.2-figure supplement 1, red lines in the first column) over a range of perceptual sensitivity is different from the cumulative distributions of sessions with confidence-related (orange) and reward expectation-related (blue) sequential effects, and is more skewed towards high motion sensitivity. We found that in monkey F, the monkey tended to have reward expectation-related sequential effects when his motion sensitivity is high (blue curve skewed towards right) and no sequential effect when his motion sensitivity is low (the blue and red distribution functions are significantly different: two-sample Kolmogorov-Smirnov test,  $p < 0.05$ ). This monkey could have confidence-related sequential effect regardless of whether his motion sensitivity is high or low (the orange distribution function is not significantly different from either the red one or the blue one: two-sample Kolmogorov-Smirnov test,  $p > 0.05$ ). For monkey C, we did not see any significant correlation between the monkey's motion sensitivity and whether the monkey had sequential effect (Figure 4.2-figure supplement 1 bottom left panel. All three distributions are the same, Kolmogorov-Smirnov test), which is different from monkey F. Other task parameters we examined, including whether being an early or late session, the ratio between large and small reward, and the correlation between reward expectation and

confidence, could not predict whether the monkeys had sequential effect or not.



**Figure 4.2-figure supplement 1. Whether a session has no sequential effect (NoSeq), confidence-related sequential effects (Conf-seq) or reward expectation-related sequential effects (RewExp-seq), cannot be predicted by the sessions' motion sensitivity, whether the session was earlier or later in data collection, ratio between large and small reward sizes, or the correlation strength between confidence and reward expectation.**

X-axes: values of the possible predictors ordered from small to large. Y-axes: empirical cumulative distribution functions of the sessions with no sequential effects, confidence-related and reward expectation-related sequential effects.

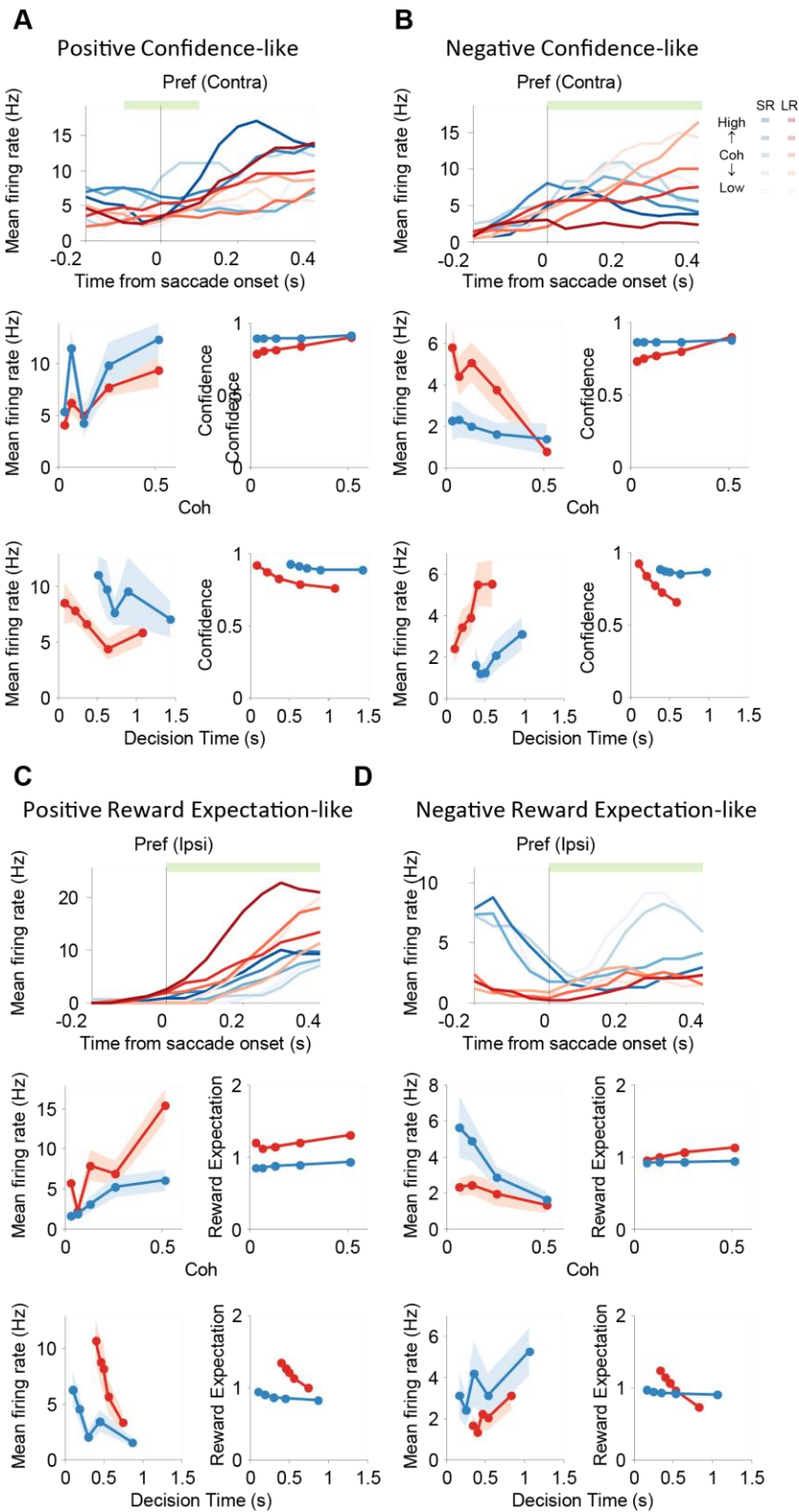
Despite the difference between the two monkeys, we found that in well-trained monkeys, confidence and reward expectation still had influence on their subsequent decision behaviors. This suggests that neural correlates of confidence and reward expectation should exist somewhere in the brain. Next we examine whether they exist in the caudate nucleus.



***Confidence and reward expectation are both represented in caudate post-decision activity.***

In chapter 3, Figure 3.3 and Figure 3.4 showed that caudate post-decision activity was modulated by motion coherence, reward size and decision time. Figure 4.1 showed that confidence and reward expectation are also jointly influenced by these three parameters. These results motivated us to consider the possibility of caudate neurons representing these two specific evaluative quantities.

Indeed, we found neural correlates of confidence and reward expectation in some caudate neurons. Figure 4.3 showed four examples: the neuron in (A) was more active when choosing small reward option, it is also more active in high coherence and short decision time trials. The neural activity patterns resemble confidence in that session. The neuron in (B) behaved almost the opposite, preferring large-reward option, low coherence and long decision time, which is similar to the negative of confidence. The neuron in (C) was more active when choosing large reward option, and preferred high coherence and short decision time, resembling the pattern of reward expectation. The neuron in (D) preferred small-reward option, low coherence and long decision time, which is similar to the negative of reward expectation in that session.

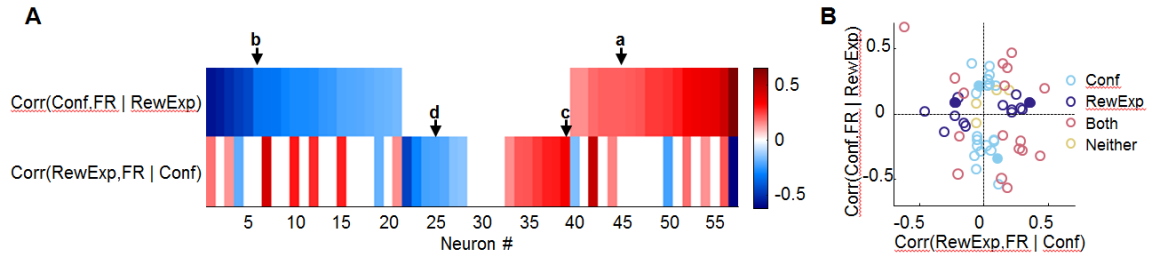


**Figure 4.3 Example post-decision caudate neural activities that resemble confidence and reward expectation.**

Activity of four example neurons that resembles positive reward expectation (A), negative reward expectation (B), positive confidence (C) and negative confidence (D). Top row: mean firing rate from 0.2s before saccade to 0.4s after saccade. Shades: coherence levels. Colors: reward size. SR: small-reward choices; LR: large reward choices. Firing rates were computed using a 200 ms running window (50-ms steps). Middle row: average firing rate (left) and confidence or reward expectation (right) as a function of coherence. Dots are the coherence levels used in that session. Ribbon: standard error. Bottom row: average firing rate (left) and confidence or reward expectation (right) as a function of decision time. Decision time was obtained from DDM fits and grouped into five quantiles. The average firing rate (dots) and standard error (ribbon) was computed from trial within each quantile. "Preferred" indicate the choice with higher firing rate on average. The time window used for computing average firing rate for middle and bottom rows are indicated by the green bars on the top row. Only correct trials were included.

To examine the prevalence of confidence and reward expectation representation among the caudate population, we used correlation analysis on the neurons' post-decision activity. Because confidence and reward expectation both varies with decision time, we first identified the neurons whose post-decision activity to the preferred direction showed decision time modulation (57 out of 142 neurons, Spearman correlation between epoch-averaged firing rate and decision time,  $p < 0.05$ ). For these neurons, we then examined if their firing rate was correlated with confidence or reward expectation. Partial-correlations were used to account for the correlation between confidence and reward expectation. We found subpopulations that represented the positive and negative values of both confidence (red and blue bars in the top row in Figure 4.4 A) and reward expectation (red and blue bars in the bottom row in Figure 4.4 A), corresponding to each of the examples shown in Figure 4.4. Moreover, confidence and reward expectation appear to be predominantly represented by distinct subpopulations of caudate neurons: only 17 neurons' activity showed correlation with both confidence and reward expectation (pink circles in Figure 4.4 B), fewer than the number of neurons whose

activity correlated with only confidence or reward expectation (36 neurons, dark and light blue circles in Figure 4.4 B). Therefore, confidence and reward expectation are both represented in individual caudate neurons.



**Figure 4.4 Confidence and reward expectation correlate with the post-saccade activity in subpopulation of caudate neurons.**

**(A)** Top row: Spearman partial correlation between confidence and post-decision neural activity in the preferred direction (FR), accounting for additional correlation with reward expectation ( $\text{Corr}(\text{FR}, \text{Conf} \mid \text{RewExp})$ ). Bottom row: Spearman partial correlation between reward expectation and post-decision neural activity in the preferred direction, accounting for additional correlation with confidence ( $\text{Corr}(\text{FR}, \text{RewExp} \mid \text{Conf})$ ). Each column in the heatmap corresponds to the same unit. Color bar: Spearman correlation coefficient (non-significant correlation coefficients ( $p \geq 0.05$ ) are plotted as white).

**(B)** Scatterplot of Spearman partial correlation coefficients for the decision time-modulated neurons. Colors indicate significant partial correlation between neural activity and confidence (light blue), reward expectation (dark blue), both (red) and neither (yellow).

Filled circles correspond to example neurons in Figure 4.3.

Although we found sessions with confidence- and reward expectation-related sequential effects, and neurons representing confidence and reward expectation, we could not find strong link between the neural activity and sequential behavior (Table 4.1): in sessions without sequential effect, we recorded both confidence-representing neurons and reward expectation-representing neurons, suggesting that the caudate nucleus encodes evaluative information even if it is not used behaviorally. In sessions with confidence-related sequential effects, we found neurons representing reward

expectation in their neural activity. Similarly, in sessions with reward expectation-related sequential effects, we found neurons whose activity represented confidence. This phenomenon is possible if confidence and reward expectation are represented simultaneously. However, this possible explanation needs to be verified using simultaneous recording in large caudate populations.

**Table 4.1. Distribution of confidence- and reward expectation-representing neurons in sessions with confidence-related sequential effects, reward expectation-related sequential effects and no sequential effects.**

	Sessions with Conf-related sequential effect	Sessions with RewExp-related sequential effect	Sessions without sequential effect
# of Neurons representing Conf	6	7	9
# of Neurons representing RewExp	7	4	4
# of Neurons representing both	2	7	3

## Discussion

Post-decision evaluation is important for learning and adaptive decision-making. By comparing the expectation and outcome, one can learn the statistical structure of the environment and the most rewarding actions and strategies. Even when a behavior is well learned, constant evaluation could help detect changes in the environment or our performance level, so that we could make necessary adjustments in time. Meanwhile, confidence could provide context in which prediction error can be appropriately interpreted. For example, a large prediction error with low confidence could be due to the task being difficult, whereas a large prediction error with high confidence might suggest changes in the environment (Purcell and Kiani, 2016)

Using a perceptual decision task that induced reward-driven biased decision behavior, we were able to partially dissociate confidence and reward expectation, two of the key evaluative quantities. We found that both confidence and reward expectation could influence subsequent decisions. Single-unit extracellular recordings showed that these two evaluative quantities were represented in the post-decision activities of subpopulation of the caudate nucleus.

Confidence and reward expectation are usually highly correlated and therefore indistinguishable in the same task. We found that the co-presence of reward magnitude asymmetry and reward-biased behavior could reduce their correlation, making them distinguishable. The key to this decorrelation is that small-reward choices tend link to high confidence, but could still lead to low reward expectation, if the reward magnitude is too low. However, to what extend are confidence and reward expectation dissociable depends on the magnitude of reward bias and the magnitude of reward asymmetry. If the difference between large- and small-reward choices is too big, or if the reward asymmetry is too small, confidence and reward expectation will still be highly correlated. One extreme scenario is when the behavior was biased by prior. For example, if the leftward motion appears more often, but the reward magnitudes of the two choices are the same, the subject might develop a prior-driven bias towards the left. In this case, his leftward choice would correspond to lower confidence, and the rightward choice would correspond to higher confidence. However, because the two choices have the same reward size, the leftward choice would also correspond to lower reward expectation, and the rightward choice would correspond to higher reward expectation. Therefore, to what extend are confidence and reward expectation correlated in the same task requires

Careful examination. Still, our task provides a paradigm in which the two quantities can be distinguished.

Our results regarding the sequential effects and neural representation of confidence and reward expectation hinge on the assumption that the confidence we computed approximates the monkeys' actual confidence. Although the confidence we computed showed patterns consistent with previously measured confidence (Kiani et al., 2014), the study could be improved by having a direct behavioral measurement of confidence or reward expectation, against which our computation could be verified. For tasks using animal subjects, it might be challenging to instruct monkeys to report their confidence on a scale (like in the human study by Kiani et al., 2014). Post-decision wagers and anticipatory licking could still be used to reflect the animal's reward expectation (Watanabe et al., 2001; Kiani and Shadlen, 2009; Fetsch et al., 2014). The setup with direct measurement would also allow us to examine whether the caudate nucleus has a causal link with confidence or reward expectation computation.

Although many studies have shown sequential effects and performance improvement based on reward prediction error, those are usually in environments with hidden structures that need to be learned by accumulating evidence across trials, or in changing environments in which the animal has to figure out when the change happens, or when the reward structure needs to be learned (Botvinick et al., 2011; Seo et al., 2012; Lak et al., 2019). Our task was designed in a way that all the information about the task is available in the current trial: the sensory information in each trial is independent from the next, and reward context changes were cued to the animals. As a result, in many sessions our monkey did not show any sequential effects. However, there could be many reasons for the presence of sequential effects in some sessions. First,

sequential effects might be hardwired into our behaviors, therefore hard to suppress, even if it is not the optimal strategy, such as confirmation bias (Talluri et al., 2018). Second, the monkeys could use past confidence and reward expectation to assess their performance and adjust their decision strategies when necessary. This is particularly possible given that the monkeys needed to calibrate sensory-encoding bias and decision-rule bias according to the reward function gradient, and they needed to know when is good enough. Finally, the monkeys' arousal level might be different from session to session. It has been shown that arousal level is linked with the balance between exploration and exploitation (Stephens and Krebs, 1986; Behrens et al., 2007), as well as the level of task engagement (Aston-Jones and Cohen, 2005). It is possible that when the animals were more alert, they were more likely to use confidence and reward expectation for exploring better strategies.

Post-decision activity in the caudate nucleus have been found to represent various kinds of monitoring- and evaluation-related information, including the value of the chosen option in non-perceptual decisions (Cromwell and Schultz, 2003; Lau and Glimcher, 2008), the difficulty level of perceptual decisions (Ding and Gold, 2010) and categorical decision boundary (Yanike and Ferrera, 2014). Our results added to the existing knowledge by showing that caudate nucleus can also carry confidence and reward expectation information in different caudate neurons. Future work with large population recording would be able to assess whether the neurons representing these two quantities coexist in two subpopulations simultaneously, or caudate neurons would represent one kind of evaluative signal at a time. If the former scenario is true, given that reward expectation is computed from confidence, it would be interesting to know if the



transformation from confidence to reward expectation is conducted by the local circuit within the caudate nucleus, or elsewhere.

Outside the caudate nucleus, the midbrain dopaminergic neurons are known to encode reward expectation before feedback and reward prediction error after feedback (Schultz, 1997; Nomoto et al., 2010; Lak et al., 2019). One study optogenetically manipulated dopamine neurons in rats during decision and during reward delivery. They found that manipulating dopaminergic neurons during decision does not influence ongoing decision or subsequent learning, whereas manipulating the neurons during feedback led to behavioral changes that could be modeled by changing the reward prediction error (Lak et al., 2019). It suggests that the reward expectation error signal might play a causal role in learning, whereas the reward expectation does not. It is possible that the reward expectation signal in the dopaminergic neurons were inherited from other brain areas, such as the caudate nucleus. In rats, a part of the striatum called striosome sends direct projections to dopaminergic neurons in the substantia nigra compacta (SNc) (Fujiyama et al., 2011; Watabe-Uchida et al., 2012). The striosome-SNc projections could carry the reward expectation signal from the striatum to SNc, which could be used for computing reward prediction error upon reward delivery.

## **Materials and Methods**

### ***Subjects***

Same as in Chapter 3. Only monkey F and monkey C were used.

### ***Behavioral task***

Same as Chapter 3.

### ***Data acquisition***

Same as Chapter 3.

### ***DDM model fitting***

Same as Chapter 2.

### ***Computation of confidence and reward expectation***

#### *Computing Confidence*

Because the motion direction and coherence in each trial was pseudo-randomly selected and unknown to the monkeys, all the information known to the monkeys at the end of a decision were their choice and decision time. Therefore, we define confidence as the estimation of their accuracy on average given the current choice and decision time, as following:

$$\text{confidence} = \begin{cases} P(\text{correct} | \text{Right choice at } T) & \text{if chosen Right} \\ P(\text{correct} | \text{Left choice at } T) & \text{if chosen Left} \end{cases} \quad \text{Eq. (1)}$$

, in which  $T$  is the decision time (reaction time minus non-decision time).

$P(\text{correct} | \text{Right/Left choice at } T)$  is computed by marginalizing over all possible coherences (this can be achieved by having performed the task over and over). For example, for rightward choices:

$$\begin{aligned}
& P(\text{correct}|\text{Right choice at } T) \\
&= \sum_{coh_i} [P(\text{correct}|\text{Right choice at } T, coh_i)P(coh_i|\text{Right choice at } T)] \\
&= \sum_{coh_i} \frac{P(\text{correct}|\text{Right choice at } T, coh_i)P(\text{Right choice at } T|coh_i)P(coh_i)}{P(\text{Right choice at } T)} \\
&= \sum_{coh_i} \frac{P(\text{correct}|\text{Right choice at } T, coh_i)P(\text{Right choice at } T|coh_i)P(coh_i)}{\sum_{coh_i} [P(\text{Right choice at } T|coh_i)P(coh_i)]}
\end{aligned} \tag{Eq. (2)}$$

, where  $coh_i$  is signed coherence (+/- for rightward and leftward motion). We defined a choice being correct as the choice being in the same direction as the motion coherence.

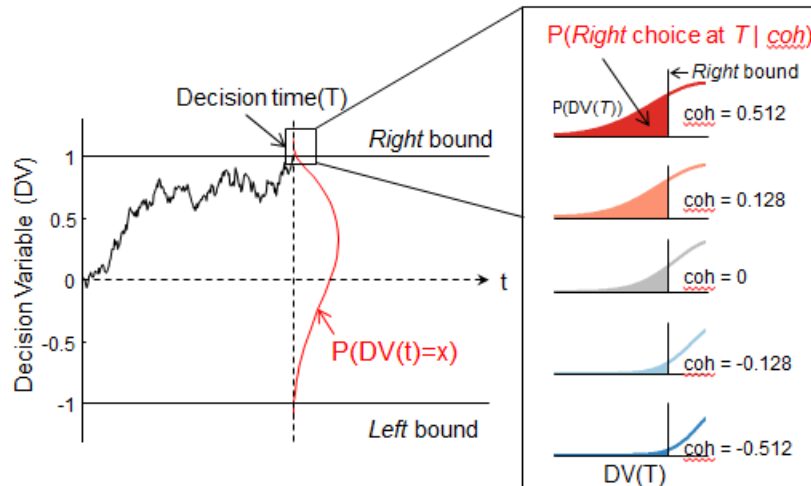
For example, for rightward choices:

$$P(\text{correct}|\text{Right choice at } T, coh_i) = \begin{cases} 1 & \text{if } coh_i > 0 \\ 0.5 & \text{if } coh_i = 0 \\ 0 & \text{if } coh_i < 0 \end{cases} \tag{Eq. (3)}$$

In our task design, each coherence had equal chance of appearance, except that  $coh=0$  happened twice as often as the other coherences:

$$P(coh) = \begin{cases} \frac{1}{\text{num. of coh}s} & \text{if } coh \neq 0 \\ \frac{2}{\text{num. of coh}s} & \text{if } coh = 0 \end{cases} \tag{Eq. (4)}$$

After plugging equation (3) and (4) into equation (2), what's left is  $P(\text{Right choice at } T|coh_i)$ . We assume that the DDM approximates the monkeys' decision-making process. This quantity was obtained by DDM simulation using the best-fitting parameters, as illustrated in Figure 4.5. For each coherence, we obtained the probability of the decision variable (DV) attaining a value  $x$  at time  $t$  ( $P(DV(t) = x)$ ), using the best fitting DDM parameters of each session and reward context. Then we computed the area underneath the probability function when  $DV > \text{Right bound}$  for rightward choices, or  $DV < \text{Left bound}$  for leftward choices.



**Figure 4.5 Related to “computing confidence” in Methods: Computing the probability of making a rightward choice at time T for a given motion coherence.**

Schematic illustrating how to compute the probability of making a rightward choice at time  $T$  for a given motion coherence ( $P(\text{Right choice at } T \mid \text{coh})$ ) using the DDM framework. For each coherence, obtain the probability of decision variable (DV) attaining value  $x$  at time  $t$  (red curve), then compute the area underneath the probability function when  $DV > \text{Right bound}$  (shaded area). For leftward choices, compute the area underneath the probability function when  $V < \text{Left bound}$ .

### Computing Reward Expectation

Reward expectation is the product of the probability of getting a reward (i.e. confidence) and the reward size associated with the choice:

$$\text{Reward Expectation} = \text{Confidence} \times \text{Reward} \quad \text{Eq. (5)}$$

, where *reward size* was set to 1 for small-reward choices and was set to the ratio between large and small reward for large-reward choices:

$$\text{Reward} = \begin{cases} \frac{\text{mean}(\text{large reward})}{\text{mean}(\text{small reward})} & \text{if large-reward choice} \\ 1 & \text{if small-reward choice} \end{cases} \quad \text{Eq. (6)}$$

### Confidence as a function of coherence

Eq. (1) shows that confidence is a function of decision time (also see Figure 4.1), which is consistent with previous study in which human subject directly reported confidence in a reaction-time task (Kiani et. al., 2014). The relationship between average confidence and coherence emerges indirectly through the relationship between coherence and decision time:

$$\begin{aligned} E(Conf|coh_i) &= \int_0^{inf} Conf P(Conf|coh_i) dConf \\ &= \int_0^{inf} Conf(T) P(T|coh_i) dT \end{aligned} \quad \text{Eq. (7)}$$

, where  $T$  is decision time;  $Conf(T)$  is confidence at decision time  $T$ ;  $P(T | coh_i)$  is the probability of making a decision at time  $T$  for a given coherence  $i$ . Simulation of the relationships between mean confidence, mean reward expectation and coherence are illustrated in Figure 4.1 J-O.

### **Neural data analysis**

For each neuron, we computed the average firing rates (FR) in a peri-saccade 300 ms window beginning at 100 ms before saccade onset and a post-saccade 400 ms window beginning at saccade onset (before reward delivery). We compared the mean firing rates across trials for the two epochs and applied further analyses on the epoch associated with the higher mean firing rate. For the chosen epoch, we compared the mean firing rate across trials for the two choice directions. The choice direction

associated with higher mean firing rate was identified as the “preferred” direction, and the opposite direction was identified as the “null” direction (see examples in Figure 4.3).

#### *Decision time modulation of the post-decision neural activity*

For each neuron, we compute the Spearman correlation between the average firing rate in the trials of the preferred direction in the chosen epoch ( $FR_{\text{pref}}$ ) and the decision time of those trials. Neurons with significant Spearman correlation coefficients ( $p < 0.05$ ) were identified as decision time-modulated neurons.

#### *Correlation between post-decision neural activity and confidence and reward expectation*

For each decision time-modulated neuron, we computed the Spearman partial correlation between the average firing rate in the trials of the preferred direction in the chosen epoch ( $FR_{\text{pref}}$ ) and the confidence and reward expectation of those trials ( $\text{corr}(FR, \text{Conf} \mid \text{RewExp})$  and  $\text{corr}(FR, \text{RewExp} \mid \text{Conf})$  in Figure 4.4), to account for the correlation between confidence and reward expectation.

#### ***Confidence-related and reward expectation-related sequential effects on the monkeys' choice and reaction time***

To examine whether confidence or reward expectation influences the monkeys' behavior in the next trial, we fit logistic function (Eq. 8) and linear function (Eq. 9) to the monkeys' choice and reaction time. Only trials after a correct trial were included.

For the choice data in each session, we fitted the following function to the choice data:

$$\log \frac{P_{\text{contra}}}{1 - P_{\text{contra}}} = (\alpha_{\text{ContraLR}} \times I_{\text{ContraLR}} + \alpha_{\text{IpsiLR}} \times I_{\text{IpsiLR}}) \times Coh$$

$$+ \beta_{\text{ContraLR}} \times I_{\text{ContraLR}} + \beta_{\text{PrevContraLR}} \times I_{\text{ContraLR}} \times Prev$$

$$+ \beta_{\text{IpsiLR}} \times I_{\text{IpsiLR}} + \beta_{\text{PrevIpsiLR}} \times I_{\text{IpsiLR}} \times Prev \quad (\text{Eq. 8})$$

, where  $P_{\text{contra}}$  is the probability of choose contralateral choice;  $Coh$  is signed coherence of current trials (+/- for motion towards contralateral/ipsilateral direction);

$$I_{\text{ContraLR}} = \begin{cases} 1 & \text{for current trial in contralateral-large reward blocks} \\ 0 & \text{other trials} \end{cases} ;$$

$$I_{\text{IpsiLR}} = \begin{cases} 1 & \text{for current trial in ipsilateral-large reward blocks} \\ 0 & \text{other trials} \end{cases} ;$$

$Prev$  is the value of the evaluative quantity (confidence or reward expectation) in the previous trials, centered to its mean across trials.

$$\text{Sequential reward-bias} = \left( \frac{\beta_{\text{PrevContraLR}}}{\alpha_{\text{ContraLR}}} - \frac{\beta_{\text{PrevIpsiLR}}}{\alpha_{\text{IpsiLR}}} \right) \times 0.5 . \text{ A positive}$$

sequential reward-bias means that when the evaluative quantity was high, the monkey biased more to the larger-reward option in the next trial.

$$\text{Sequential choice-bias} = \left( \frac{\beta_{\text{PrevContraLR}}}{\alpha_{\text{ContraLR}}} + \frac{\beta_{\text{PrevIpsiLR}}}{\alpha_{\text{IpsiLR}}} \right) \times 0.5 . \text{ A positive}$$

sequential choice-bias means that when the evaluative quantity was high, the monkey biased more to the contralateral option in the next trial.

For the RT data in each session, we fitted the following function to the trials when both the previous and the current trial were correct:

$$RT = (a_0 + a_{\text{Choice}} \times I_{\text{Choice}} + a_{\text{RewSize}} \times I_{\text{RewSize}} + a_{\text{ChoiceRew}} \times I_{\text{Choice}} \times I_{\text{RewSize}}) \times Coh$$

$$+ b_0 + b_{\text{Choice}} \times I_{\text{Choice}} + b_{\text{RewSize}} \times I_{\text{RewSize}} + b_{\text{ChoiceRew}} \times I_{\text{Choice}} \times I_{\text{RewSize}}$$

$$+ b_{\text{Prev}} \times Prev + b_{\text{PrevRew}} \times I_{\text{RewSize}} \times Prev \quad (\text{Eq. 9})$$

, in which  $Coh$  is the un-signed motion coherence in the current trials (positive for both directions)

$$I_{Choice} = \begin{cases} 1 & \text{for current contralateral choices} \\ -1 & \text{for current ipsilateral choices} \end{cases} ;$$

$$I_{RewSize} = \begin{cases} 1 & \text{if chose large reward in current trial} \\ -1 & \text{if chose small reward in current trial} \end{cases} ;$$

$Prev$  is defined the same way as in the logistic function.

$b_{Prev}$  is the reward size-independent sequential effect on RT. This term being positive/negative means that when the evaluative quantity was high, the monkey tended to speed up/slowdown in the next trial.

$b_{PrevRew}$  is the reward size-dependent sequential effect on RT. This term being positive/negative suggests that when the evaluative quantity was high, the monkeys tended to speed up/slowdown when choosing the large-reward option and slowdown/speed up when choosing the small-reward option in the next trial.

Log likelihood of a model is the sum of the log likelihoods of the logistic and linear fits. For the confidence version (Conf-seq) and the reward expectation version (RewExp-seq), we fitted the following three models:

(1) "Full model": sequential effect included in both the logistic function and the linear function (i.e.  $\beta_{PrevContraLR}$ ,  $\beta_{PrevContraLR}$ ,  $b_{Prev}$  and  $b_{PrevRew}$  are all included);

(2) "Choice-only": sequential effect included in the logistic function but no sequential effect in the linear function (i.e.  $b_{Prev}$  and  $b_{PrevRew}$  are not included);

(3) "RT-only": no sequential effect in the logistic function but sequential effect included in the linear function (i.e.  $\beta_{PrevContraLR}$  and  $\beta_{PrevContraLR}$  are not included).



We also fitted a no-sequential effect model (NoSeq)—none of  $\beta_{PrevContraLR}$ ,  $\beta_{PrevContraLR}$ ,  $b_{Prev}$  or  $b_{PrevRew}$  is included.

AIC was used for model comparison. For Conf-seq and RewExp-seq, among the three models, we selected the one that has the smallest AIC as the best-fitting model. Then we compared the AICs between the best-fitting models in the Conf-seq version and the RewExp-seq version, together with the NoSeq model. The model with the smallest AIC was used to interpret whether the sequential effects in a session were confidence-related or reward expectation related, or did not exist.

## Reference

- Aston-Jones G, Cohen JD (2005) An Integrative Theory Of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annu Rev Neurosci* 28:403–450.
- Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214–1221.
- Botvinick MM, Niv Y, Barto AG (2011) Hierarchically organised behaviour and its neural foundations: A reinforcement-learning perspective. In: *Modelling Natural Action Selection*, pp 264–299. Cambridge University Press.
- Cromwell HC, Schultz W (2003) Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J Neurophysiol* 89: 2823-2838.
- Daw ND, Doya K (2006) The computational neurobiology of learning and reward. *Curr Opin Neurobiol* 16:199-204.
- Ding L, Gold JI (2010) Caudate encodes multiple computations for perceptual decisions. *J Neurosci* 30:15747–15759 .
- Fetsch CR, Kiani R, Newsome WT, Shadlen MN (2014) Effects of Cortical Microstimulation on Confidence in a Perceptual Decision. *Neuron* 83:797–804.
- Fujiyama F, Sohn J, Nakano T, Furuta T, Nakamura KC, Matsuda W, Kaneko T (2011) Exclusive and common targets of neostriatofugal projections of rat striosome neurons: A single neuron-tracing study using a viral vector. *Eur J Neurosci* 33:668-677.
- Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455:227–231.
- Kiani R, Corthell L, Shadlen MN (2014) Choice certainty is informed by both evidence

- and decision time. *Neuron* 84:1329–1342.
- Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324:759–764.
- Lak A, Okun M, Moss M, Gurnani H, Farrell K, Wells MJ, Reddy CB, Kepecs A, Harris KD, Carandini M (2019) Neural basis of learning guided by sensory confidence and reward value. *bioRxiv*:411413.
- Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58:451–463.
- Mathys C, Daunizeau J, Friston KJ, Stephan KE (2011) A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* 5:39.
- Meder D, Kolling N, Verhagen L, Wittmann MK, Scholl J, Madsen KH, Hulme OJ, Behrens TEJ, Rushworth MFS (2017) Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. *Nat Commun* 8:1942.
- Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasley B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci* 15:1040–1046.
- Nassar MR, Wilson RC, Heasley B, Gold JI (2010) An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci* 30:12366–12378.
- Nomoto K, Schultz W, Watanabe T, Sakagami M (2010) Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. *J Neurosci* 30:10692–10702.
- Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology

- of value-based decision making. *Nat Rev Neurosci* 9:545–556.
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340.
- Schultz W (1997) A Neural Substrate of Prediction and Reward. *Science* 275:1593–1599.
- Schultz W (2015) Neuronal Reward and Decision Signals: From Theories to Data. *Physiol Rev* 95:853–951.
- Seo M, Lee E, Averbeck BB (2012) Action Selection and Action Value in Frontal-Striatal Circuits. *Neuron* 74:947–960.
- Stephens DW, Krebs JR (John R. (1986) *Foraging theory*. Princeton University Press.
- Sutton RS, Barto AG (1998) *Reinforcement learning : an introduction*. MIT Press.
- Talluri BC, Urai AE, Tsetsos K, Usher M, Donner TH (2018) Confirmation Bias through Selective Overweighting of Choice-Consistent Evidence. *Curr Biol* 28:3128-3135.e8
- Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N (2012) Whole-Brain Mapping of Direct Inputs to Midbrain Dopamine Neurons. *Neuron* 74:858-873.
- Watanabe M, Cromwell HC, Tremblay L, Hollerman JR, Hikosaka K, Schultz W (2001) Behavioral reactions reflecting differential reward expectations in monkeys. *Exp Brain Res* 140:511–518.
- Yanike M, Ferrera VP (2014) Interpretive monitoring in the caudate nucleus. *Elife* 3:1–16
- Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. *Neuron* 46: 681-692.

## CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

Yunshu Fan, Joshua I Gold, Long Ding

Using a task that encouraged monkeys to combine sensory and reward information for decision-making, I found that the way the monkeys combined sensory and reward information generally conformed to a drift-diffusion model (DDM). However, the specific biasing strategies they used were suboptimal, and varied from session to session and monkey to monkey. By linking the monkeys' strategies with their individual reward functions, we found that the suboptimal and variable strategies were consistent with a common rational heuristic. This heuristic is sensitive to the individual variabilities of the reward functions across monkeys and sessions, which led to the individual variations in the monkeys' idiosyncratic biasing strategies.

By recording in the caudate nucleus while the monkeys were performing the reward-biased perceptual decision task, we found that the caudate nucleus represented information related to the decision process throughout the trial. Specifically, before the decision starts, some caudate neurons represented reward context, which could be used to establish the reward bias towards a specific option later in the trial, similar to a starting value bias in the DDM. During decision formation, both sensory and reward information were combined in a subpopulation of individual caudate neurons. This result, together with a subsequent study that established the causal role of caudate nucleus in combining sensory and reward information using electrical micro-stimulation, suggests that caudate neurons may participate in combining sensory and reward information for decision formation. After decision, we found that sensory evidence and reward

information continued to be represented in individual caudate neurons, but not all of them conform to an intuitive “reward expectation” signal. We further found that our task design allowed us to disambiguate reward expectation from confidence; these two quantities were usually indistinguishable in conventional task designs (Kepecs et al., 2008; Lak et al., 2019). This allowed us to find out that while some caudate neurons’ post-decision activity represented reward expectation, some other caudate neurons’ activities represented confidence. We also found that confidence and reward expectation each influenced monkeys’ decision behaviors in the future in a subset of sessions, suggesting that the confidence-like and reward expectation-like signals encoded in the caudate post-decision activity could be used for evaluation.

These findings open up a number of future directions as follow.

## **Experimental/Task design**

### ***Importance of carefully designed complex behavior tasks***

Our results highlight how, in the context of studying complex behavior, the task design can reveal aspects of behavior that are otherwise hidden. In the context of goal-directed behavior, the brain can combine multiple sources of information adaptively in response to changes in the environment, as well as to changes in internal states. Internal states might refer not only to the preference for a specific reward, but also to the proficiency to make accurate perceptual judgements. In many asymmetric-reward experimental paradigms, the ability to adapt to reward preference leads us to observe reward-driven bias in the behavior. Our task added an additional manipulation, i.e., either large or small reward was given only when the perception was correct. This tapped into the brain’s ability to adapt the reward-driven bias to the proficiency of making

accurate perceptual judgements. This design also allowed us to observe that the monkeys calibrated their biasing strategy with regard to their motion sensitivity: when motion sensitivity was high, i.e., when the monkeys were able to make more accurate perceptual decisions, they tended to have less reward bias and their strategies were closer to the optimal. In contrast, when motion sensitivity was low, i.e., when the monkeys were making less accurate perceptual decisions, they tended to have more reward bias, and their strategies were farther from the optimal. This can be understood from the reward function's perspective. When motion sensitivity is high, the peak of the reward function is closer to no bias, and the plateau of the reward function is also smaller. This will encourage sub-optimal but good enough strategies to be closer to the peak, which is also closer to no bias. On the contrary, when motion sensitivity is low, the peak of the reward function corresponds to large bias. Meanwhile, the plateau of the reward function is big. This will allow more deviation from the optimal strategy to be good enough, which magnifies the magnitude of bias. This bias-sensitivity tradeoff not only explained the individual variability among our monkeys' decision behaviors, but could also be one of the reasons why many previous studies using similar tasks found that subjects appeared to adopt different decision strategies (Voss et al., 2004; Bogacz et al., 2006; Simen et al., 2009; Summerfield and Koechlin, 2010; Leite and Ratcliff, 2011; Mulder et al., 2012; Goldfarb et al., 2014; Cicmil et al., 2015). We were able to uncover this relationship between bias and proficiency due to the specific complexity in our task design.

The decisions we make every day are usually complex and are influenced by many factors. While simpler tasks are useful in probing the underlying neural mechanisms, complex tasks can help discover the effect of some factors that are only

revealed during complex behaviors. Our results highlight the importance of using a carefully designed complex behavioral task with systematic quantitative modeling and analyses to understand various factors that influence adaptive behaviors and strategies.

***Task design should allow key variables to be dissociable.***

When some variables are involved in generating the behavior but are not directly accessible via measurements, computational modeling is often used to extract these latent variables for further hypothesis testing and interpreting neural computations. We need to make sure that the task design will allow different latent variables in the model to be distinguishable from one another. For example, we were able to examine whether the reward biased the monkeys' sensory-encoding (*me*-bias) or decision rule-setting (*z*-bias) via DDM fitting, because our task provided reaction time data. The reaction time data is crucial because, in DDM, *me* and *z* could generate similar choice biases, but the RT distributions they generated are qualitatively different, especially when compared between error and correct trials (Figure 2.3-figure supplement 1). Similarly, we were able to examine the neural correlates of confidence and reward expectation in the same task because our task and the reward-biased behavior it induced allow the two variables to exhibit different patterns (Figure 4.1). In contrast, many previous studies that use equal-reward task design were not able to identify if the behavioral effects and neural correlates they were studying were related to confidence or reward expectation (Kepecs et al., 2008; Lak et al., 2019). If distinguishing two variables is the key to the scientific question under study, the behavior task needs to be designed (or redesigned) so that the two variables generate qualitatively different behavioral readouts.



### ***Task design is an iterative process.***

As a research project develops, preliminary results can inform how a task should be modified. My thesis project has shown me that it is hard to predict where the data might lead us. Sometimes we might obtain results that were not expected, leading us to new analyses to better understand the data. I will highlight this through three examples. First, when studying the biasing behavior of the monkeys, we did not start with investigating the specific heuristic. This came up because it appeared to better account for our monkeys' behavior patterns. Because our task was not ideally suited to observing the "gradient ascent" searching process, we could not know when and how fast this process happens. Answers to these questions require modification to our experiments, such as: (1) collecting data during learning, and (2) removing the cue of the reward context, and making the reward context switching unpredictable so as to increase the chance of observing the "gradient ascent" learning and the adjusting process in the behavior. A second example relates to a prediction of the rational, satisficing heuristic. It predicts that, if the motion sensitivity is too low, the reward function gradient might be steeper along the  $z$  dimension than along the  $me$  dimension. This would lead to a rational suboptimal decision strategy that overly biases  $z$  to the adaptive direction and compensate with  $me$  biased to the small reward direction. This hypothesis cannot be tested with our data, because all of the monkeys used in our study have gone through years of training on the motion discrimination task. In theory, they belong to the expert group whose reward functions all favor the overly biasing  $me$  strategy. Whether a subject with low motion sensitivity tends to favor an overly biasing  $z$  strategy would be better tested in subjects with less training. Finally, Chapter 4 focused on two specific evaluative signals—confidence and reward expectation, an angle that we did not plan

when we designed the task. Therefore our task did not include a behavioral report of the monkeys' confidence levels. Although our task allowed confidence and reward expectation to be distinguishable based on the way we computed them, our results regarding the neural representation of confidence and reward expectation could have been much stronger if we had a direct confidence measurement. Adding the behavioral report would also allow us to examine if changes in the neural correlates results in changes in confidence, or if manipulating the neural activities that represent confidence and reward expectation would lead to corresponding changes in the behavior reports (in the spirit of SENSE AND THE SINGLE NEURON: Probing the Physiology of Perception (Parker and Newsome, 1998)). These examples illustrate how the task and experiments should be adjusted dynamically in the light of new, especially unplanned, results.

### **Caudate nucleus and reward-biased perceptual decision-making**

#### ***Distinct role of the caudate nucleus in combining sensory and reward information.***

We discovered that during decision-formation period, both sensory and reward information are represented in individual caudate neurons, but not in the format of a decision variable (DV) in the DDM, especially in terms of bound crossing. A subsequent study using micro-stimulation in the caudate nucleus during decision making in the same task confirmed that the caudate nucleus is causally involved in combining sensory and reward information (Doi et al., 2019). Meanwhile, cortical neurons in LIP have been found to correlate with a DV that combines sensory evidence and non-sensory reward bias and prior bias (Rorie et al., 2010; Hanks et al., 2011). These results suggest that caudate nucleus is an intermediate station where sensory and reward information are combined, playing a modulatory role, which could feed to the final DV elsewhere in the

brain. It is still unclear what exact computations are performed by the caudate nucleus and whether and how they contribute to the DV formation. Previously, computational models based on the specific anatomical structure of the basal ganglia have been developed for action selection, decision-making and reinforcement learning (Redgrave et al., 1999; Bogacz and Gurney, 2007; Samejima and Doya, 2007; Hikosaka et al., 2014; Caballero et al., 2018). Adding reward-biasing mechanisms to these models could serve as a starting point for understanding the computations performed in the caudate nucleus and generally in the basal ganglia. Many of these models involve distinct computations in the direct, indirect and hyperdirect pathways and the neural plasticity modulated by different dopamine receptors. Given that we also observed diverse patterns of caudate neural activity in single-unit recording, it is very likely that different caudate neurons might be involved in different pathways or computational units. Future experiments using large scale recording, the ability to identify the pathway they are in, and the neuronal type (at least in terms of D1 or D2 receptor expression), will allow for a better understanding of the specific computations performed by the caudate nucleus within the basal ganglia circuitry.

### ***Caudate nucleus in the context of basal ganglia circuitry***

Meanwhile, the caudate nucleus is only the input station of the interconnected basal ganglia circuitry. Information has to go through multiple stages of processing via not only different pathways, but also recurrent loops, before sending out to other brain areas. Even though we did not observe bound crossing-like activity patterns in the caudate nucleus, a decision variable can be formed in downstream areas. Alternatively, downstream areas might further modify the information they receive from the caudate

nucleus, making the relationship between specific activity in caudate neurons and how sensory and reward information are combined behaviorally more complicated. Isolating the computational role of caudate nucleus might be similar to looking at one part of a very complicated mathematical solution. Recording in caudate and downstream areas simultaneously will allow us to understand how sensory and reward information are combined in the basal ganglia as a whole and interpret the computational role of each individual nucleus within the larger circuit.

### ***Diverse computations in the caudate nucleus and information flow***

We found that the caudate nucleus represented diverse computational quantities before, during and after making a decision. It would be interesting to know whether these diverse quantities are sent out separately to distinct targets, or whether they are all sent to a range of target regions. For example, decision formation-related information might be projected to motor-related areas, such as LIP, FEF and SC (Horwitz and Newsome, 1999; Roitman and Shadlen, 2002; Ding and Gold, 2012), for execution, whereas evaluation-related information might be projected to areas involved in evaluation, error signal encoding and metacognition, such as midbrain dopaminergic neurons, anterior and posterior cingulate cortex and medial prefrontal cortex (Schultz, 1997; Behrens et al., 2007; Matsumoto et al., 2007; Heilbronner and Platt, 2013). Hypotheses like this need to be verified by recording in multiple brain areas simultaneously during the same task.

## **New theoretical frameworks**

A key foundation of our study is the theoretical framework, i.e., DDM. Even though it might not be implemented in the brain on a physical level, as defined by David Marr (Marr and Poggio, 1977), computational models as such still provide us a useful angle to examine behavior, neural activity and the links between them. In chapter 2, the DDM helped us discover the specific deviation pattern of the monkey and the optimal strategy and the link between bias and sensitivity. In chapter 3, it prompted us to understand the information representation from the perspective of biases (in terms of time-independent  $z$ -like, or time-increasing  $me$ -like) and decision formation (in terms of decision variable). In chapter 4, it provided the method for computing confidence. However, our results also pointed out the need for new theoretical frameworks to be developed in many aspects as discussed below.

## ***Frameworks and tools to understand individual variability***

Individual variability commonly exists, although it might not be commonly reported. Yet, it might sometimes reflect common factors that influence behavior. In chapter 2, through examining the biasing strategy in the context of reward function, we discovered one plausible mechanism for the individual variability we observed in our monkeys – idiosyncratic perceptual sensitivity leads to different adjustments of biases in response. However, we still don't know how general this principle is in other kinds of behavioral paradigms. It is very likely that this is only one mechanism underlying individual variability. At least one other source of individual difference could come from the differences in the mental complexity among subjects, both in terms of model complexity (Tavoni et al., 2019) and how much past experience is used to inform future

actions (Glaze et al., 2018). Currently there is no unified framework/guideline to systematically examine individual variability.

### ***The interaction between evaluation and adaptive decision-making***

Even though we showed that the confidence and reward expectation influenced subsequent decision behaviors (Chapter 4), our result was mainly descriptive. Through what exact computation do they exert their evaluative role is still unclear. Many previous studies on the neural correlates of evaluative signal also largely stayed on the level of post-decision neurons representing task-relevant information, without going into how the signals were used for the specific behavior. Evaluative signal in terms of reward prediction error has been studied in the context of reinforcement learning (Sutton and Barto, 1998; Doya, 2007). It has also been applied to link evaluative neural signals with behavior in learning paradigms (Lak et al., 2019). In such a setting, the goal of evaluation is to update the values in order to figure out the option with the best value. However, outside learning, the goal of adaptive decision behaviors might not be reward-maximization, yet evaluation might still be needed for minor strategy adjustments. In this case, we need new theoretical frameworks for specifying what evaluation is needed and how it could be used for behavioral adjustments.

### **Towards computational psychiatry**

Computational modeling can link behavior with underlying neural mechanisms. The recently emerged field of computational psychiatry is trying to apply the insights and methodology from computational modeling to investigate the links between circuit impairments and psychiatric symptoms. This might enhance our understanding of the

psychiatric disorders not only on the level of molecular and physiological features (such as *neurexins* mutation in Autism patients (Ching et al., 2010), and hyper-excitability in epilepsy patients (Scharfman, 2007)), but also on the level of circuit functions (Wang and Krystal, 2014). This could also lead to behavioral diagnosis for circuit dysfunction, more effective targeting of the impaired circuit during treatment, and the use of behavioral biomarkers for symptom monitoring during and after treatment. Psychiatric disorders usually involve complex behaviors. My study described in the thesis assessed such complex behaviors. Specifically, our results could open up a new dimension for assessing behavior, i.e. the ability to adapt our strategy to our internal proficiency and accuracy in performing a task. Even though we did not find the location(s) in the brain that link strategy with proficiency, it is possible that damage to such brain structures could impair the subjects' ability to adjust strategies effectively when his/her proficiency changes.

## Reference

- Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214–1221.
- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113:700–765.
- Bogacz R, Gurney K (2007) The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput* 19:442–477.
- Caballero JA, Humphries MD, Gurney KN (2018) A probabilistic, distributed, recursive mechanism for decision-making in the brain. *PLoS Comput Biol* 14: e1006033.
- Ching MSL et al. (2010) Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet Part B Neuropsychiatr Genet* 153B: 937-947.
- Cicmil N, Cumming BG, Parker AJ, Krug K (2015) Reward modulates the effect of visual cortical microstimulation on perceptual decisions. *eLife* 4:e07832.
- Ding L, Gold JI (2012) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cereb Cortex* 22:1052–1067.
- Doi T, Fan Y, Gold JI, Ding L (2019) The caudate nucleus controls coordinated patterns of adaptive, context-dependent adjustments to complex decisions. *bioRxiv*:568733.
- Doya K (2007) Reinforcement learning: Computational theory and biological mechanisms. *HFSP J* 1:30.
- Glaze CM, Filipowicz ALS, Kable JW, Balasubramanian V, Gold JI (2018) A bias–variance trade-off governs individual differences in on-line learning in an



- unpredictable environment. *Nat Hum Behav* 2:213–224.
- Goldfarb S, Leonard NE, Simen P, Caicedo-Núñez CH, Holmes P (2014) A comparative study of drift diffusion and linear ballistic accumulator models in a reward maximization perceptual choice task. *Front Neurosci* 8:148.
- Hanks TD, Mazurek ME, Kiani R, Hopp E, Shadlen MN (2011) Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J Neurosci* 31:6339–6352.
- Heilbronner SR, Platt ML (2013) Causal evidence of performance monitoring by neurons in posterior cingulate cortex during learning. *Neuron* 80: 1384-1391.
- Hikosaka O, Kim HF, Yasuda M, Yamamoto S (2014) Basal Ganglia Circuits for Reward Value–Guided Behavior. *Annu Rev Neurosci* 37:289–306.
- Horwitz GD, Newsome WT (1999) Separate signals for target selection and movement specification in the superior colliculus. *Science* 284:1158-1161.
- Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455:227–231.
- KN G, MD H, P R (2015) A new framework for cortico-striatal plasticity: behavioural theory meets in vitro data at the reinforcement-action interface. *PLOS Biology* 13: e1002034.
- Lak A, Okun M, Moss M, Gurnani H, Farrell K, Wells MJ, Reddy CB, Kepecs A, Harris KD, Carandini M (2019) Neural basis of learning guided by sensory confidence and reward value. *bioRxiv*:411413.
- Leite FP, Ratcliff R (2011) What cognitive processes drive response biases? A diffusion model analysis. *Judgm Decis Mak* 6:651–687.
- Marr DC, Poggio T (1976) From understanding computation to understanding neural

- circuitry. Artificial Intelligence Laboratory. A.I. Memo. Massachusetts Institute of Technology. AIM-357.
- Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. *Nat Neurosci* 10:647–656.
- Mulder MJ, Wagenmakers E-J, Ratcliff R, Boekel W, Forstmann BU (2012) Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J Neurosci* 32:2335–2343.
- Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci* 21:227–277.
- Redgrave P, Prescott TJ, Gurney K (1999) The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience* 89:1009–1023.
- Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22:9475–9489.
- Rorie AE, Gao J, McClelland JL, Newsome WT (2010) Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS One* 5:e9308.
- Samejima K, Doya K (2007) Multiple representations of belief states and action values in corticobasal ganglia loops. *Ann N Y Acad Sci* 1104:213–228.
- Scharfman HE (2007) The neurobiology of epilepsy. *Curr Neurol Neurosci Rep* 7:348–354.
- Schultz W (1997) A Neural Substrate of Prediction and Reward. *Science* 275:1593–1599.
- Simen P, Contreras D, Buck C, Hu P, Holmes P, Cohen JD (2009) Reward Rate

- Optimization in Two-Alternative Decision Making: Empirical Tests of Theoretical Predictions. *J Exp Psychol Hum Percept Perform* 35:1865–1897.
- Summerfield C, Koechlin E (2010) Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. *Front Hum Neurosci* 4:208.
- Sutton RS, Barto AG (1998) Reinforcement learning : an introduction. MIT Press.
- Tavoni G, Balasubramanian V, Gold JI (2019) The complexity dividend: when sophisticated inference matters. *bioRxiv*:563346.
- Voss A, Rothermund K, Voss J (2004) Interpreting the parameters of the diffusion model: an empirical validation. *Mem Cognit* 32:1206–1220.
- Wang X-J, Krystal JH (2014) Neuron Perspective Computational Psychiatry. *Neuron* 84: 638-654.