

**A value-based explanation for lapses in  
perceptual decisions**

Sashank PISUPATI

*A thesis submitted in fulfillment of the requirements  
for the degree of Doctor of Philosophy*

Watson School of Biological Sciences  
Cold Spring Harbor Laboratory

April 22, 2019

*“Not all who wander are lost.”*

J.R.R. Tolkein

**Cold Spring Harbor Laboratory  
Watson School of Biological Sciences**

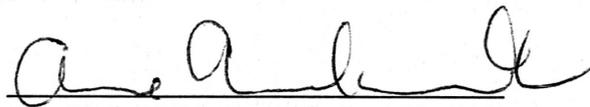
*Sashank Pisupati*

We, the dissertation committee for the above candidate for the  
Doctor of Philosophy degree,  
hereby recommend the acceptance of this dissertation.



**Bo Li, Professor**

Watson School of Biological Sciences  
Cold Spring Harbor Laboratory  
*Chair*



**Anne Churchland, Associate Professor**

Watson School of Biological Sciences  
Cold Spring Harbor Laboratory  
*Research Mentor*



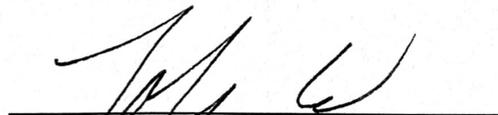
**Stephen Shea, Associate Professor**

Stony Brook University  
Stony Brook, NY  
*Academic Mentor*



**Anthony Zador, Professor**

Watson School of Biological Sciences  
Cold Spring Harbor Laboratory  
*Committee Member*



**Joshua Gold, Professor**

University of Pennsylvania School of Medicine  
Philadelphia, PA  
*External Examiner*

This dissertation is accepted by the Watson School of Biological Sciences.



**Alexander Gann, Dean**  
Watson School of Biological Sciences



**Date**

## *Acknowledgements*

I would like to thank my supervisor, Dr. Anne Churchland for her constant guidance, mentorship and support throughout my time as a graduate student. I'd also like to thank the members of my thesis committee, Tony Zador, Steve Shea and Bo Li, for their valuable critique. I am extremely grateful to my wonderful and talented experimental collaborator Lital Chartarifsky, and to our resident rodent whisperer Anup Khanal, without whom this project would not have been possible. I would also like to thank Barry Burbach, Angela Licata, Steven Gluf, Liete Einhorn, Dennis Maharjan, Edward Lu and Alexa Pagliaro for their technical assistance. I would like to thank Matt Kaufman, Simon Musall, Onyekachi Odoemene, Ashley Juavinett, Farzaneh Najafi, Akihiro Funamizu, Priyanka Gupta, Anne Urai, James Roach, Colin Stoneking and Tatiana Engel, for their invaluable scientific feedback. I'd like to thank Partha Mitra, Alexander Tolpygo and Stephen Savoia for help with slicing and imaging virus injected brains. I am eternally grateful to Rob Phillips for inspiring in me a faith in theory. I want to thank my parents for their support and kind words of encouragement, even from afar. Finally, my partner in science and in life, Diksha Gupta, for her brilliant scientific critique, love and unwavering faith in me.

# Contents

<b>Acknowledgements</b>	<b>ii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Ideal observer framework . . . . .	2
1.1.1 Signal detection theory . . . . .	2
1.1.2 Bayesian Decision theory . . . . .	2
1.1.3 Optimal multisensory integration . . . . .	3
1.2 Thesis outline . . . . .	5
<b>2 Deviations from optimal decision-making in rodents</b>	<b>6</b>
2.1 Biases . . . . .	8
2.1.1 Biases arising from weighing irrelevant features . . . . .	8
2.2 Deviations from optimal integration . . . . .	10
2.3 Lapses . . . . .	11
2.4 Discussion . . . . .	13
<b>3 Traditional models of lapses in perceptual decisions</b>	<b>14</b>
3.1 Motor error model . . . . .	14
3.2 Inattention model . . . . .	16
3.3 Discussion . . . . .	18
<b>4 Lapses reflect exploration, reveal subjective value</b>	<b>19</b>

4.1	Value uncertainty in perceptual decision making . . . . .	19
4.1.1	Exploration-exploitation trade-off . . . . .	20
4.1.2	Effects of perceptual uncertainty on exploration . . . . .	22
4.1.3	Model comparison . . . . .	22
4.2	Testing the predictions of the exploration model . . . . .	22
4.2.1	Manipulating reward magnitude . . . . .	24
4.2.2	Manipulating reward probability . . . . .	25
4.2.3	Effects of uncertainty . . . . .	25
4.2.4	Effects of reward history . . . . .	26
4.2.5	Effects of contingency reversal . . . . .	28
4.3	Discussion . . . . .	29
<b>5</b>	<b>Leveraging lapses to infer the stage of decision-making deficits</b>	<b>30</b>
5.1	Informativeness of lapses in perturbation experiments . . . . .	30
5.2	Identifying candidate multisensory areas . . . . .	30
5.3	Prefrontal, striatal inactivations resemble value deficit . . . . .	32
5.3.1	Asymmetric effects on lapses . . . . .	34
5.3.2	Model comparison . . . . .	34
5.3.3	Sensory and motor controls . . . . .	34
5.3.4	Discussion . . . . .	36
<b>6</b>	<b>Conclusion and Perspectives</b>	<b>37</b>
6.1	Summary of findings . . . . .	37
6.2	The influence of value on perceptual decisions . . . . .	38
6.3	Insights into individual differences . . . . .	39
6.4	Insights into the role of prefrontal cortex, striatum . . . . .	39
6.5	Uncertainty-guided exploration and dopamine . . . . .	40

<b>A</b>	<b>Methods</b>	<b>41</b>
A.1	Behavioral training . . . . .	41
A.2	Behavioral manipulations . . . . .	42
A.2.1	Congruence manipulation . . . . .	42
A.2.2	Reward manipulation . . . . .	43
A.3	Tracing experiments . . . . .	43
A.4	Neural manipulations . . . . .	44
A.4.1	Inactivation with muscimol . . . . .	45
A.4.2	Histology . . . . .	45
A.5	Behavioral analysis . . . . .	46
A.5.1	Psychometric curves . . . . .	46
A.6	Modeling . . . . .	46
A.6.1	Ideal observer model . . . . .	46
A.6.2	Inattention model . . . . .	50
A.6.3	Motor error model . . . . .	51
A.6.4	Exploration model . . . . .	52
A.6.5	Fitting and comparisons . . . . .	53
<b>B</b>	<b>Supplementary figures</b>	<b>54</b>
	<b>Bibliography</b>	<b>60</b>

# List of Figures

1.1	Optimal multisensory integration . . . . .	4
2.1	Multisensory task in rats . . . . .	7
2.2	Hybrid readout predicts biased multisensory criterion . . . . .	9
2.3	Deviations from optimal predictions . . . . .	10
2.4	Accounting for lapses improves fit, restores optimality . . . . .	11
2.5	Lapses are reduced on multisensory decisions . . . . .	12
3.1	Traditional models of lapses . . . . .	15
3.2	Lapses are increased on neutral trials . . . . .	17
4.1	Explore-exploit trade-off in perceptual decisions . . . . .	21
4.2	Perceptual uncertainty increases exploration . . . . .	23
4.3	Predictions for reward manipulations . . . . .	24
4.4	Reward manipulations confirm exploration model predictions . . . . .	27
4.5	Effects of learning on lapses . . . . .	28
5.1	Lapses are informative about decision deficits . . . . .	31
5.2	Identifying candidate multisensory areas . . . . .	32
5.3	Prefrontal, striatal inactivations resemble value deficit . . . . .	33
5.4	Inactivation effects are not sensory or motor . . . . .	35
B.1	Predictions from models of lapses . . . . .	55

B.2	Alternative models of lapses . . . . .	56
B.3	Individual fits . . . . .	57
B.4	Contingencies and sides . . . . .	58
B.5	No effects of movement parameters . . . . .	59

# List of Abbreviations

<b>A1</b>	Primary auditory cortex
<b>AAV</b>	Adeno Associated Virus
<b>AIC</b>	Akaike Information Criterion
<b>BDT</b>	Bayesian Decision Theory
<b>BIC</b>	Bayes Information Criterion
<b>GABA</b>	Gamma Amino-Butyric Acid
<b>GFP</b>	Green F luorescent protein
<b>M2</b>	Secondary motor cortex
<b>MAP</b>	Maximum A Posteriori
<b>pStr</b>	Posterior Striatum
<b>SDT</b>	Signal Detection Theory
<b>V1</b>	Primary visual cortex

## List of Symbols

$\mu$	Mu, criterion or midpoint of psychometric	Hz
$\sigma$	Sigma, noise standard deviation or inverse slope of psychometric	Hz
$\gamma$	Gamma, Lower asymptote of psychometric	Proportion
$\lambda$	Lambda, Upper asymptote of psychometric	Proportion
$\epsilon$	Epsilon, proportion of trials on which random choice is made	Proportion
$\beta$	Beta, Inverse temperature or inverse exploratoriness	$\mu\text{L}^{-1}$
$c$	True stimulus category	unitless
$s$	True stimulus rate	Hz
$a$	Correct choice	unitless
$\hat{a}$	Subject's choice	unitless
$x_a$	Noisy auditory observation of rate	Hz
$x_v$	Noisy visual observation of rate	Hz
$r_L$	Leftward reward magnitude	$\mu\text{L}$
$r_R$	Rightward reward magnitude	$\mu\text{L}$
$Q_L$	Leftward action value	$\mu\text{L}$
$Q_R$	Rightward action value	$\mu\text{L}$

## Chapter 1

# Introduction

Perceptual decisions are decisions that involve categorizing sensory observations into one of many categories, in order to pick an appropriate course of action. For instance, a faint movement or rustling in the bushes might be a harmless breeze or a dangerous predator, two stimuli that warrant very different responses. These decisions are often hindered by noise in the sensory observations - which might arise from nuisance variables (eg. occlusions in the environment, irrelevant sounds in the background) or from imperfect measurement (eg. noise in sensory transduction), leading to uncertainty in subjects' beliefs about the true category of the stimulus, and giving rise to errors in their decisions. Further complicating the matter is the fact that multiple sensory modalities might be informative about the decision, but could be corrupted by different amounts of noise (eg. Low-light conditions might make vision less reliable than hearing), so it might be desirable to differentially weigh these while making a decision. Integrating noisy perceptual information in order to make accurate decisions in the face of uncertainty is critical to organisms' survival, and understanding how this is achieved in the brain is a long-standing question in systems neuroscience (Hanks and Summerfield, 2017).

## 1.1 Ideal observer framework

A long-standing tradition in the study of perceptual decisions is the use of ideal observer models to define theoretical upper limits on the performance of an observer as a function of the noise in their observations. Such models are normative, since they define what an animal *should* do in order to make the best use of the limited information and neural resources available to it. Moreover, they offer a reference to which animal behavior can be compared, allowing one to identify computational constraints in the decision making process (Gershman, Horvitz, and Tenenbaum, 2015; Lieder and Griffiths, 2019).

### 1.1.1 Signal detection theory

One of the earliest examples of this is Signal Detection Theory (Green and Swets, 1966), which posits that observers receive a noisy measurement of the perceptual signal (an “internal response”), and compare it to an internal criterion in order to make a decision about whether they detected the signal or not. This predicts that the overall rate of errors in the subjects’ decisions (i.e. misses and false alarms) should reflect the overlap in the distributions of internal responses with and without the signal (The discriminability index or  $d'$ ) independent of their criterion, revealing the level of noise in their observations - higher levels of noise would give rise to an increase in both kinds of errors. On the other hand, changes in criterion should change the relative amounts of these errors - a higher criterion would lead to fewer false alarms but more misses for the same level of noise.

### 1.1.2 Bayesian Decision theory

While signal detection theory explains how performance is bounded by observation noise, it makes no prescriptions about what criterion a subject should choose, rather emphasizing

the trade-off between different kinds of errors. Bayesian decision theory (Dayan and Daw, 2008) extends this approach by explicitly incorporating the costs of different kinds of errors as well as any prior knowledge, to prescribe a single criterion that the subject should use, in order to maximize the average or *expected* reward from their decisions. In this theory, subjects combine the *Likelihood* that a noisy observation was generated by one or another stimulus category with a *Prior* belief about the probability of occurrence of these stimulus categories in the world, to yield a *Posterior* belief about the category, given the observations. The posterior belief is combined with learnt knowledge about the costs and benefits of performing different actions, to compute expected value or expected utility which can then be maximized. Bayesian decision theory explicitly takes uncertainty into account, predicting that subjects should rely more on their priors when their sensory observations are uncertain (Stocker and Simoncelli, 2006), and should weigh information from different sensory modalities in inverse proportion to their respective uncertainties (Ernst and Bulthoff, 2004).

### 1.1.3 Optimal multisensory integration

Decisions based on multiple sensory modalities provide a particularly unique opportunity to test the predictions of Bayesian decision theory (BDT), since it predicts the optimal performance for multisensory stimuli as a function of the noise in the individual sensory modalities, which can be empirically estimated through the use of unisensory stimuli (Figure 1.1). Taking this approach, a number of studies have shown that humans (Ernst and Bulthoff, 2004) and non-human primates (Gu, Angelaki, and DeAngelis, 2008; Hou et al., 2018) both integrate multisensory information near-optimally. This approach has recently been extended to rodents (Raposo et al., 2012; Nikbakht et al., 2018), a powerful model system that enables high throughput behavior (Carandini and Churchland, 2013; Burgess et al., 2017; Odoemene et al., 2018) and fine-scale perturbation of brain regions. However,

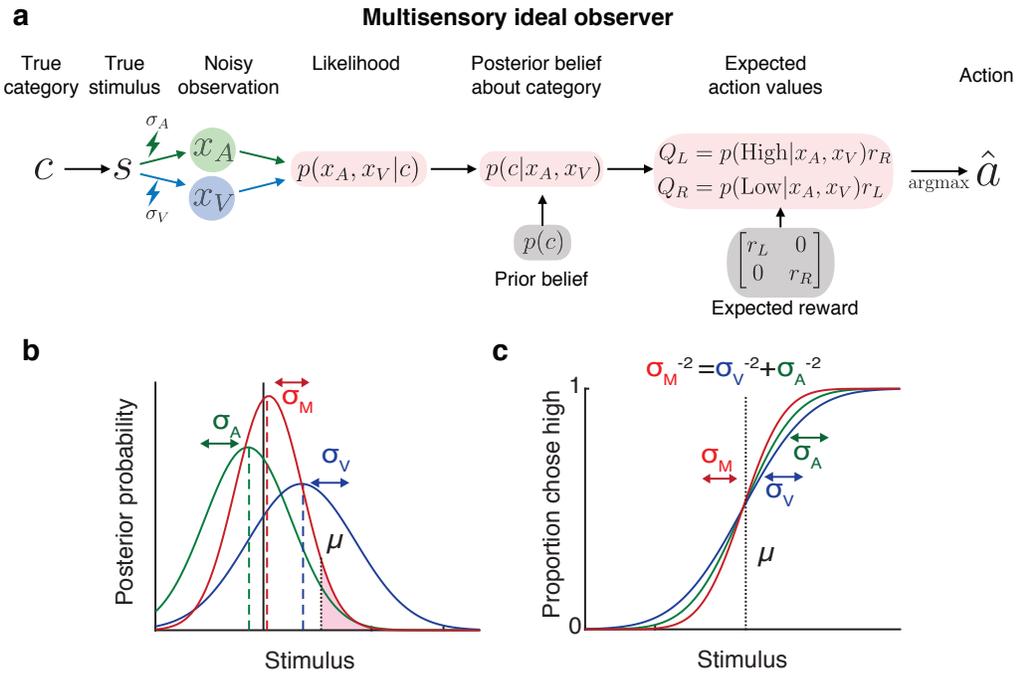


FIGURE 1.1: **Optimal multisensory integration.** (a) Multisensory ideal observer constructed from BDT, for stimuli that could belong to "High" or "Low" categories. (b) Posterior beliefs based on unisensory observations (blue, green) or optimal multisensory integration (red) with uncertainties denoted by  $\sigma$ , and optimal decision criterion denoted by  $\mu$  (dotted black). (c) Predictions for behavior on unisensory and multisensory trials.

a number of recent studies have reported significant violations of the optimal upper bound on performance (Raposo et al., 2012; Nikbakht et al., 2018; Hou et al., 2018), calling into question the assumptions of the Bayesian model. One possible explanation for this is there are imprecisions or suboptimalities in the decision making process that are unaccounted for by standard Bayesian Decision Theory, giving rise to apparent supra-optimality (Beck et al., 2012; Drugowitsch et al., 2016; Shalom and Zaidel, 2018; Noel, 2018; Rahnev and Denison, 2018). Another possibility is that these deviations reflect a form of "bounded" rationality i.e. the optimal use of limited resources (Lieder and Griffiths, 2019), and are strategic rather than accidental.

## 1.2 Thesis outline

This thesis aims to provide a formal treatment of deviations from optimality in perceptual decision-making, and reconsider these deviations from the perspective of value-based decision making, ultimately providing a synthesis of the two. In Chapter 2 I describe deviations from optimal multisensory decisions in rats, focusing in particular on “lapses”, an often ignored form of error that seems to plague many perceptual studies. In Chapter 3 I outline the traditional models of lapses, and use behavioral manipulations to demonstrate their insufficiency. In Chapter 4, I propose that lapses can instead arise deliberately from trying to balance an exploration-exploitation trade-off, and test the predictions of this model by manipulating value. Finally in Chapter 5, I discuss how lapses can be informative about deficits in decision making, and leverage them to reconcile the role of prefrontal cortex and striatum in perceptual and value-based decisions.

## Chapter 2

# Deviations from optimal decision-making in rodents

One of the reasons for the success of ideal observer models is their ability to parsimoniously capture behavior in humans and non-human primates with remarkable accuracy. This is evident in a number of studies in humans and non-human primates. At the same time, a number of other studies in non-human primates and rodents have revealed various deviations from ideal observer predictions.

To investigate this further, we trained a large cohort of rats on an established multisensory decision-making task (Raposo et al., 2012; Raposo, Kaufman, and Churchland, 2014; Sheppard, Raposo, and Churchland, 2013; Licata et al., 2017), in which freely moving rats judge whether the fluctuating rate of a 1000 ms series of auditory clicks and/or visual flashes (rate range: 9 - 16 Hz) is high or low compared with an abstract category boundary. Rats are trained to report their decisions by poking into a right or left port for high or low decisions respectively, for which they are rewarded with a small drop of water (Figure 2.1).

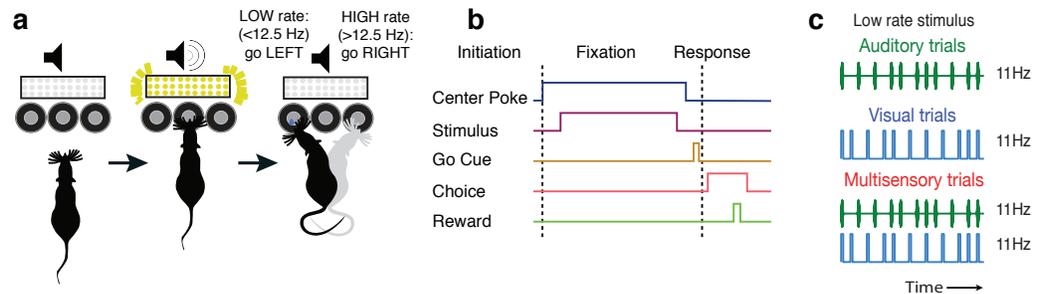


FIGURE 2.1: **Multisensory task in rats.** (a) Schematic of task showing rats initiating trials (left), fixating while a stimulus is played (center) and reporting choices with orienting movements (right). (b) Timeline of task events. (c) Example low rate stimulus on auditory (top), visual (center) and multisensory (bottom) trials.

We assessed performance using a psychometric curve, i.e., the probability of high-rate decisions as a function of stimulus rate. The ideal observer model makes predictions about three key aspects of the psychometric curve: The midpoint, the slope and the asymptotes. If the animal had the correct priors i.e. knew that high and low rate stimuli were equally likely, and had correctly learnt the category boundary and expected rewards, then the ideal observer model predicts that the animal's criterion should match the true category boundary of 12.5 Hz. It also predicts that the slope of the unisensory psychometric curves should reflect the noise in each of these modalities: the higher the standard deviation ( $\sigma$ ) of sensory noise, the more uncertain the animal's estimate of the rate and the shallower the psychometric curve. On multisensory trials, the ideal observer model predicts that the animal should integrate its visual and auditory observations, yielding a more accurate estimate of rate and driving a steeper psychometric curve. It also predicts the maximum possible slope the animal could achieve from optimal integration, as a function of the unisensory noise. Finally, it predicts that the animal should make no errors for stimuli far from the criterion, and hence the psychometric curve should asymptote at 0 and 1. Hence, we fit the data separately on the three conditions, to measure sensory noise and criterion and compare these to ideal observer predictions.

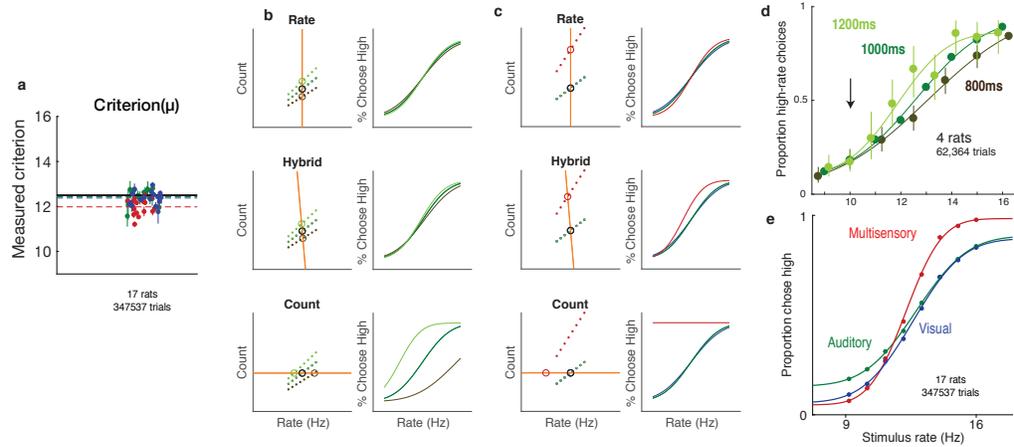
## 2.1 Biases

On unisensory trials, the animals were unbiased on average, and had midpoints indistinguishable from the true criterion of 12.5, however they showed a small but significant high rate bias (0.5 Hz) away from the true criterion on multisensory trials (Figure 2.2). In practice, the midpoint of the psychometric curve could be biased away from the true category boundary for a number of reasons - incorrectly learnt likelihoods, incorrect priors or unequal subjective utilities from the two actions. Here we consider another reason for seemingly biased behavior that may actually stem from a subject optimally learning likelihoods and rewards, but doing so over “irrelevant” stimulus features that happen to be correlated with “task-relevant” ones (Odoemene et al., 2018).

### 2.1.1 Biases arising from weighing irrelevant features

Unlike the ideal observer, the subject does not have perfect *a priori* knowledge about the task-relevant stimulus features or the category boundary, and can only learn these by keeping track of which stimulus features were followed by rewarded leftward or rewarded rightward actions. This corresponds to drawing a classification boundary in the space of all possible stimulus features between stimuli with different experienced rewards, and the projections of this boundary on any particular stimulus feature determines its weight on the decision.

For multisensory decisions, since the rate in both modalities is relevant and correlated (but corrupted by independent noise), this corresponds to equal weights on the two modalities, scaled by the respective noise variances, yielding the familiar uncertainty-weighted decision rule. However, if some stimulus feature that was supposed to be “irrelevant” happened to be correlated with the relevant feature of rate, from the subjects’ point of



**FIGURE 2.2: Hybrid readout predicts biased multisensory criterion.**

(a) Fitted criterions of 17 rats showing slight bias on multisensory trials (red). (b) Predictions of rate (top), count (bottom) and hybrid (middle) readouts for different durations of stimulus (light green: 1200ms, dark green: 1000ms, brown: 800 ms) (c) Predictions of the same decoders on multisensory trials. (d) Performance on duration manipulations, showing slight bias as predicted by hybrid readout. (e) Performance on multisensory trials (red) showing slight bias as predicted by hybrid readout.

view it would be informative about reward and hence relevant, and should ideally receive a non-zero weight.

In our task, the rate is correlated with the total count of events, raising the possibility that the animal might use the count instead of rate (or some combination of the two). We tested this by changing the duration of the stimulus on unisensory (auditory) trials, maintaining the same rate but increasing or decreasing the count. This yielded minimal but noticeable biases that could be captured by a hybrid readout that largely weighed rate, but slightly weighed count (15 times less than rate). The same hybrid readout predicted a slight high-rate bias in the animal's criterion on multisensory trials (Figure 2.2), since the total count is doubled on multisensory trials compared to unisensory trials with corresponding rates, which we observed

## 2.2 Deviations from optimal integration

The ideal observer model predicts that the squared inverse slope on multisensory trials should equal the harmonic mean of the squared inverse slopes on unisensory trials, if the animal integrates optimally across modalities. While most studies of multisensory integration have reported near-optimal performance, a few studies have reported better-than-optimal performance in rodents (Raposo et al., 2012; Nikbakht et al., 2018) and primates (Hou et al., 2018). This violates the upper bound predicted by the ideal observer model, since performance cannot be better than optimal by definition, suggesting that one or more of the model's assumptions must be incorrect. To confirm this finding, I compared predicted  $\sigma$  extracted from unisensory trials with the measured  $\sigma$  on multisensory trials. Indeed, the measured multisensory  $\sigma$  was significantly better than the optimal prediction from unisensory trials ( $p=0.0012$ ), replicating previous findings (Figure 2.3).

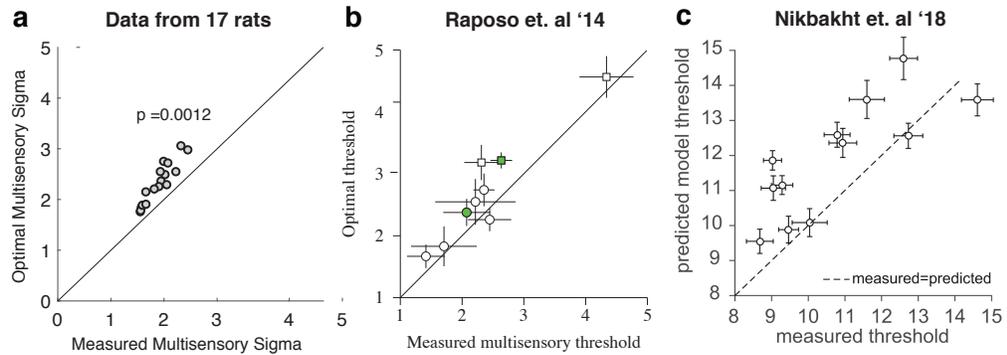


FIGURE 2.3: **Deviations from optimal predictions.** (a) Data from 17 rats showing significantly better performance than the optimal prediction (points are above unity line, paired t-test,  $p=0.0012$ ), replicating previous reports from (a) rats (squares) and humans (circles) on the same audio-visual task (Reproduced from Raposo et al Nat. Neuro '19) and (b) rats on a visuo-tactile task (Reproduced from Nikbakht et al Neuron '18)

## 2.3 Lapses

The last key prediction of the ideal observer model is that performance should asymptote at 0 and 1 for very easy stimuli. However, in practice this is often not the case, as has been observed in a large number of psychophysical studies in humans (Treutwein and Strasburger, 1999; Klein, 2001; Wichmann and Hill, 2001; Flesch et al., 2018; Mihali et al., 2018), non-human primates (Law and Gold, 2009; Cohen and Maunsell, 2009; Cloherty et al., 2019), mice (Busse et al., 2011; Burgess et al., 2017; Odoemene et al., 2018; Pinto et al., 2018; Wang et al., 2018; Lak et al., 2018) and rats (Raposo et al., 2012; Brunton, Botvinick, and Brody, 2013; Scott et al., 2015; Nikbakht et al., 2018; Piet et al., 2017; Mendonca et al., 2018). These deviations in the asymptotes are known as lapses, since they are assumed to arise from occasional lapses in attention, or errors in motor execution (Wichmann and Hill, 2001). To account for lapses, I allowed the asymptotes to deviate from 0 and 1, by adding two additional parameters per condition (Figure 2.4).

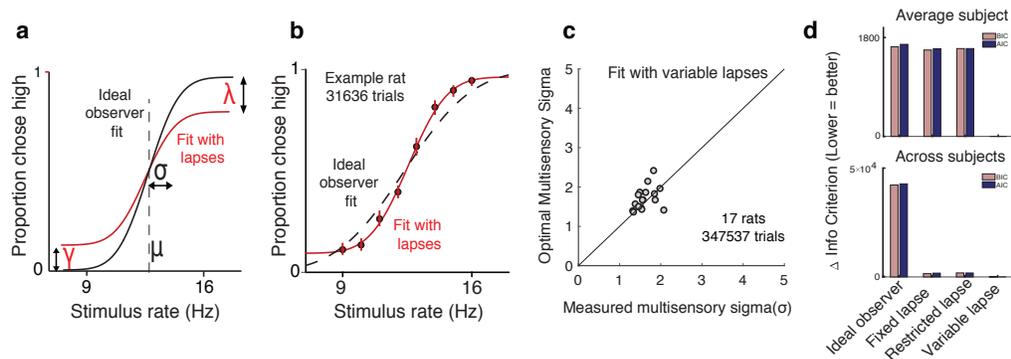
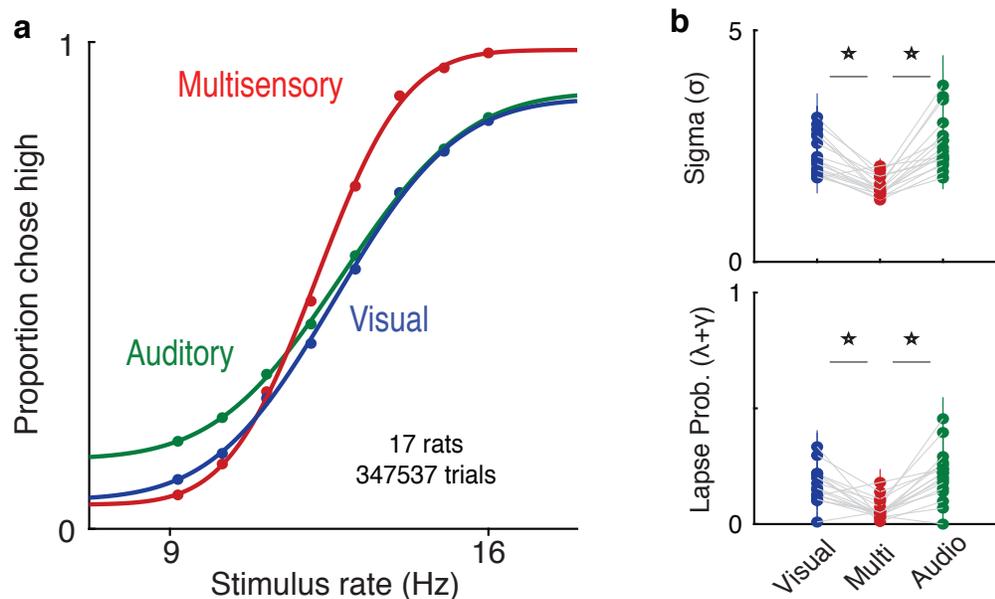


FIGURE 2.4: **Accounting for lapses improves fit, restores optimality.** (a) Schematic of ideal observer model (black) and model with lapses (red). (b) Example rat data showing improved fit with lapses. (c) Optimality is restored when lapses are accounted for. (d) Model comparison showing that the best fitting model allows lapses to vary across conditions ("Variable lapses")

This yielded much better fits to the data, and interestingly the integration was no

longer distinguishable from the optimal prediction, suggesting that the supra-optimality arose from biases in our estimates of  $\sigma$  as a result of improperly handling lapses. Indeed, forcing the lapse rate to be less than 0.1 or forcing it to be fixed across conditions (as is traditional in many studies - see Wichmann and Hill, 2001; Dobs, Ma, and Reddy, 2017; Mihali et al., 2018) also yielded worse fits and significant supra-optimality ( $p = 0.0003$ ,  $0.0018$  respectively). Such biases in slope estimates are well documented in the literature (Prins, 2012; Prins and Kingdom, 2018), and emphasize the importance of properly handling lapses.

Interestingly, we observed that animals had a lower lapse probability on multisensory trials than on unisensory trials (Figure 2.5). This was consistently observed across animals (0.06 on average for multisensory trials, compared to 0.17 for visual,  $p=1.4e-4$  and 0.21 for auditory,  $p=1.5e-5$ ).



**FIGURE 2.5: Lapses are reduced on multisensory decisions.** (a) Average behavior of 17 rats on visual (blue), auditory (green) and multisensory (red) trials. (b) Fitted values of sigma ( $\sigma$ , top) and lapse parameters ( $\gamma + \lambda$ , bottom) showing a significant reduction in both parameters (paired t-test) on multisensory trials

A lower rate of lapses on multisensory trials has also been reported on a visual-tactile task in rats (Nikbakht et al., 2018) and a vestibular integration task in humans (Bertolini et al., 2015) and can potentially account for previous reports of apparently supra-optimal integration (Nikbakht et al., 2018; Hou et al., 2018; Raposo et al., 2012).

## 2.4 Discussion

The focus of this chapter has been on accounting for deviations from various predictions of the ideal observer. Rats seem to display an apparent supra-optimality in multisensory integration, however this disappears once we account for an apparent sub-optimality - errors at the asymptotes of the psychometric curve or lapses. While the addition of lapse parameters captures the data well, their source remains unclear, as well as the puzzling observation that they are reduced on multisensory trials, a performance enhancement not predicted by the ideal observer.

## Chapter 3

# Traditional models of lapses in perceptual decisions

A number of possible sources of noise have been proposed to explain lapses, typically adding a fixed amount of error peripheral to the decision-making process (Figure 3.1). These fall into two major classes - post-decisional noise typically thought of as motor errors, and pre-decisional noise thought of as fluctuations in attention. Here I describe these two models and test whether they can account for the lapses we observe in behavior.

### 3.1 Motor error model

One class of explanations for lapses relies on a fixed amount of noise added after a decision has been made, commonly referred to as “post-categorization” noise (Erlich et al., 2015) or “decision noise” (Law and Gold, 2009). Such noise could arise from errors in motor execution (e.g. “finger errors” , Wichmann and Hill, 2001), non-stationarities in the decision rule arising from computational imprecision (Findling et al., 2018), suboptimal weighting of choice or outcome history (Roy et al., 2018; Busse et al., 2011) or random variability added for the purpose of exploration (eg. “ $\epsilon$ -greedy” decision rules). However,

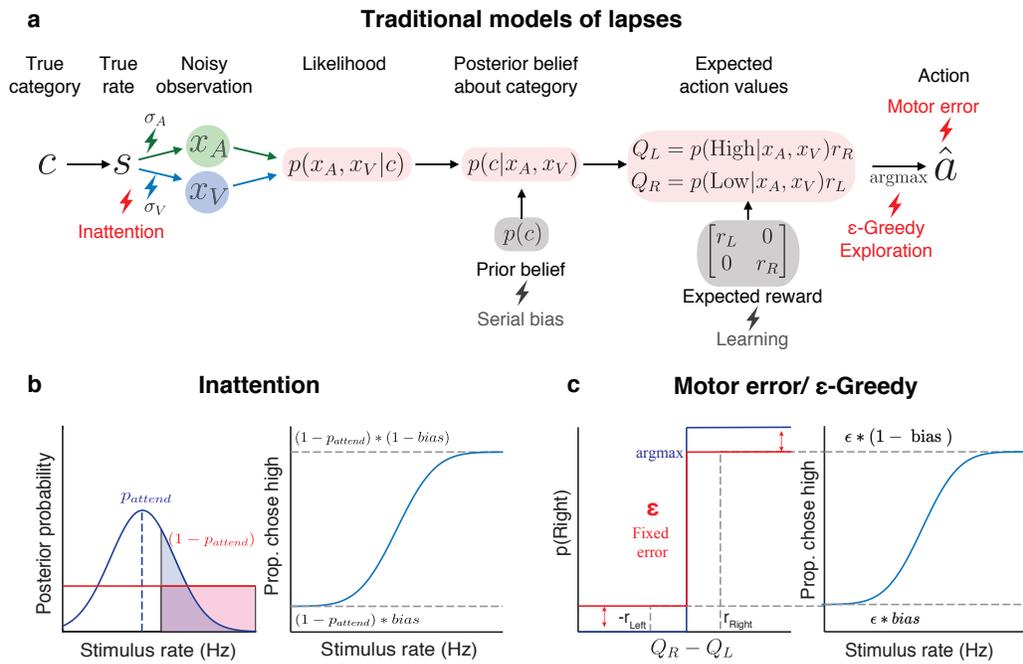


FIGURE 3.1: **Traditional models of lapses.** (a) Schematic showing different sources of noise traditionally proposed to give rise to lapses (Red thunderbolts) or exacerbate them (grey thunderbolts). (b) Inattention model of lapses (c) Motor error model of lapses

these sources should hinder decisions equally across conditions, predicting a fixed lapse rate, which cannot explain our observation of condition-dependent lapse rates.

### 3.2 Inattention model

Another class of explanations for lapses relies on pre-decision noise added due to fluctuating attention, which is often operationalized as a small fraction of trials on which the subject fails to attend to the stimulus (Wichmann and Hill, 2001). On these trials, it is assumed that the subject cannot specify the stimulus (i.e. sensory noise with infinite variance, Bays, Catalao, and Husain, 2009) and hence guesses randomly or in proportion to prior beliefs. This model can be thought of as a limiting case of the Variable Precision model, which assumes that fluctuating attention has a more graded effect of scaling the sensory noise variance (Garrido, Dolan, and Sahani, 2011), giving rise to heavy tailed estimate distributions, resembling lapses in the limit of high variability (Shen and Ma, 2019; Zhou et al., 2018). Temporal forms of inattention have also been proposed to give rise to lapses, where the animal ignores early or late parts of the evidence (impulsive or leaky integration, Erlich et al., 2015).

If lapses occur due to inattention, then the sum of the two lapse rates should reflect the probability of not attending, so the lower lapse probability on multisensory trials could arise from increased attention on those trials, perhaps due to their increased bottom-up salience (i.e. two streams of stimuli instead of one). To test this, we leveraged a multisensory condition that manipulates uncertainty without changing salience (Raposo et al., 2012). Specifically, we interleaved standard matched-rate multisensory trials with “neutral” multisensory trials for which the rate of the auditory stimuli ranged between 9-16 Hz, while the visual stimuli was always 12 Hz. This rate was so close to the category boundary (12.5 Hz) that it did not provide compelling evidence for one choice or the other, thus reducing

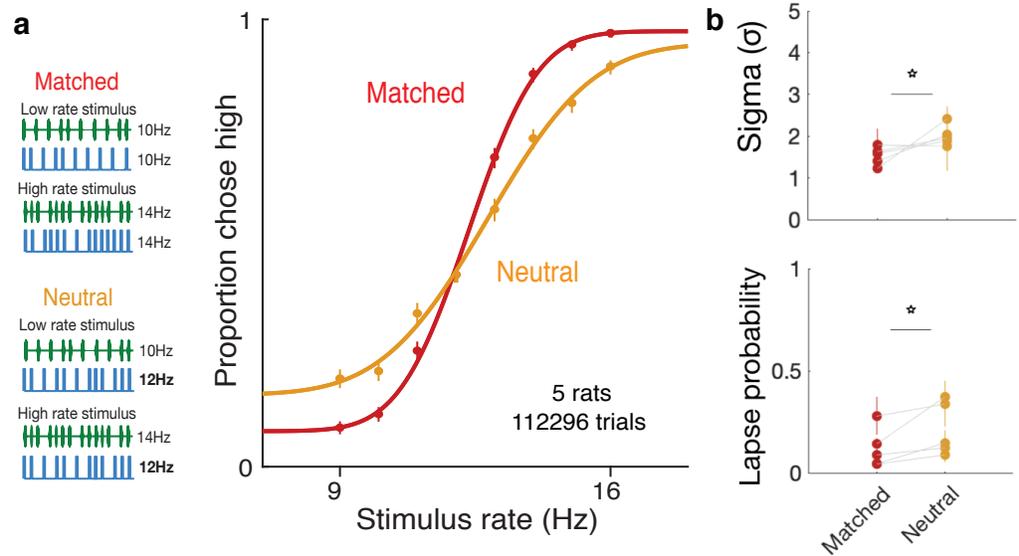


FIGURE 3.2: **Lapses are increased on neutral trials.**(a) Schematic of a low rate stimulus with matched visual and auditory rates (top left) or with uninformative visual stimuli (neutral; bottom left) and average behavior of 5 rats on matched (red) and neutral (orange) trials. (b) Fitted values of sigma ( $\sigma$ , top) and lapse parameters ( $\gamma + \lambda$ , bottom) showing a significant increase in both parameters (paired t-test) on neutral trials

the information in the multisensory stimulus and increasing uncertainty. However, since both “neutral” and “matched” conditions are multisensory, they should be equally salient, and since they are interleaved, the animal would be unable to identify the condition without actually attending to the stimulus. According to the inattention model, matched and neutral trials should have the same rate of lapses, only differing in their  $\sigma$ .

Contrary to this prediction, we observed higher lapse rates on “neutral” trials, where the uncertainty was high, than on “matched” trials, where the uncertainty was lower (Figure 3.2). The dependence of lapses on uncertainty is reminiscent of the dependence of lapse on uncertainty observed when comparing unisensory vs. multisensory trials

### **3.3 Discussion**

Together, our results suggest that lapses do not simply arise from stimulus independent noise added peripherally to the decision making process, but depend critically on the stimulus condition, suggesting that they may reflect an integral part of the decision-making process. Condition-dependent lapses have been previously observed both across modalities (Raposo et al., 2012; Bertolini et al., 2015; Nikbakht et al., 2018) and within modalities (Mihali et al., 2018; Nunez et al., 2018). Moreover, lapses have been shown to change over time within the same condition (Law and Gold, 2009) suggesting that they may reflect task-specific, associative processes.

## Chapter 4

# Lapses reflect exploration, reveal subjective value

### 4.1 Value uncertainty in perceptual decision making

Having observed that traditional explanations of lapse fail to account for the behavioral observations, I re-examined one of the key assumptions made when applying Bayesian Decision Theory to perceptual decision - that the subject has full knowledge of the expected rewards (Dayan and Daw, 2008). This assumption is often made since most perceptual decision making tasks have a fixed, deterministic reward structure and subjects' performance is typically measured after extensive training.

However, in general this assumption may not hold true, either because the subject is still in the process of learning the expected rewards (Law and Gold, 2009), or because they forget the expected rewards over time (Gershman, 2015), or because they incorrectly assume non-stationarity in their environments (Yu and Cohen, 2009).

In any of these situations, the subject may have uncertainty about the expected rewards, in which case choosing the action with the highest expected value, or "exploiting",

is no longer the optimal policy. In such situations, it is advantageous to “explore”, or occasionally pick an action whose expected reward one is uncertain about. Moreover, the subject can treat this uncertainty in a Bayesian manner across trials, and update its beliefs across trials as it receives rewards. We can simulate this in an agent with a two-stage Bayesian update - within every trial, the agent accumulates evidence in order to infer the stimulus category, and across trials, it updates its beliefs about the expected rewards based on the outcomes of previous trials. The uncertainty in its expected rewards depends on its policy as well as the uncertainty in the stimulus category of previous trials.

#### 4.1.1 Exploration-exploitation trade-off

The effect of the agent’s policy on uncertainty defines an exploration-exploitation trade-off in its decisions: Random exploration would equally reduce uncertainty about all combinations of expected reward (High->Left, High->Right, Low->Left, Low->Right), but would yield reward at chance. On the other hand, a greedy policy (i.e. always exploiting) would maximize rewards, but only reduce uncertainty about the higher value stimulus-action pairs (High->Right, Low->Left) leaving the expected rewards on the lower value stimulus-action pairs uncertain.

One way to balance this trade-off is to explore in proportion to one’s uncertainty - exploring more when one is more uncertain about expected rewards. This form of uncertainty-guided exploration is known as Thompson sampling (Gershman, 2018) and has been shown to be asymptotically optimal in partially observable environments (Leike et al., 2016). Thompson sampling can be implemented through a sampling scheme - where the agent samples a set of expected rewards from its beliefs and then maximizes with respect to the sample. This scheme reduces to a dynamic softmax decision rule in the case of gaussian beliefs about expected rewards (Gershman, 2018), a decision rule that is commonly used in value-based decision making owing to its tunable  $\beta$  parameter (Figure 4.1).

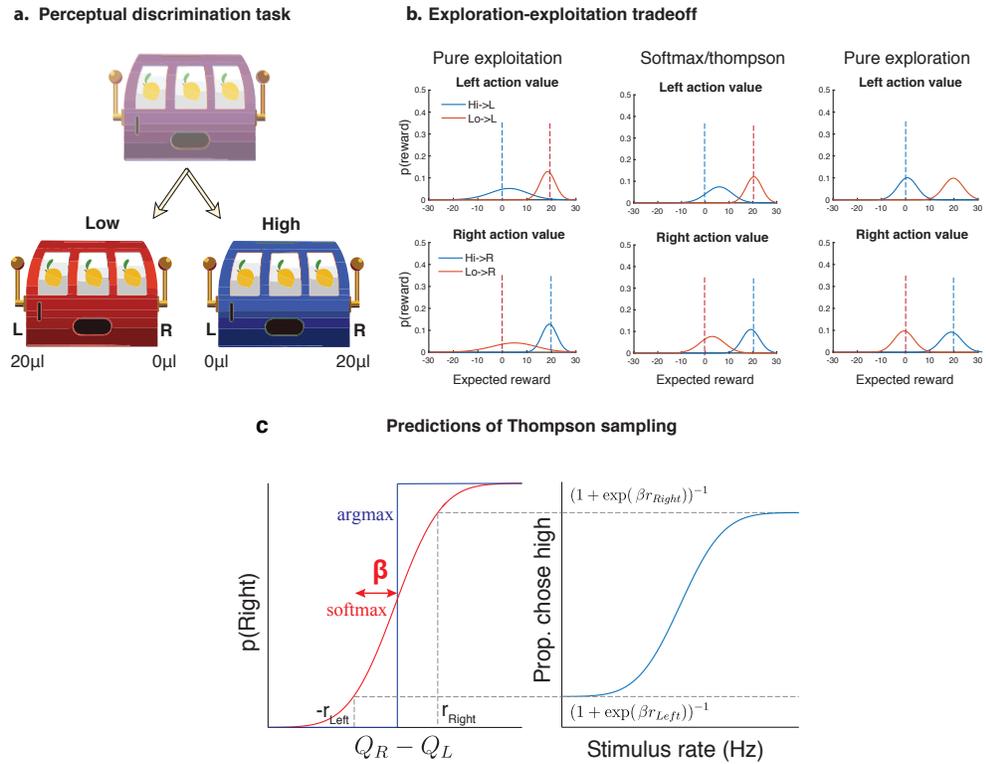


FIGURE 4.1: **Explore-exploit trade-off in perceptual decisions** (a) Conceptualizing perceptual decisions as a partially observable contextual bandit - on any trial the animal infers the probability that the stimulus was high (blue) or low (red), and chooses the arm (Right or Left) with highest expected reward. (b) Simulation of a Bayesian learner’s beliefs about reward. If it always exploits (left), it gets rewarded often but is more uncertain about lower-reward arms. If it always explores (right), it is equally certain about all rewards but only gets rewarded at chance. Thompson sampling (center) balances these two opposing objectives. (c) Thompson sampling reduces to a softmax decision rule with an uncertainty-dependent  $\beta$ , giving rise to lapses

### 4.1.2 Effects of perceptual uncertainty on exploration

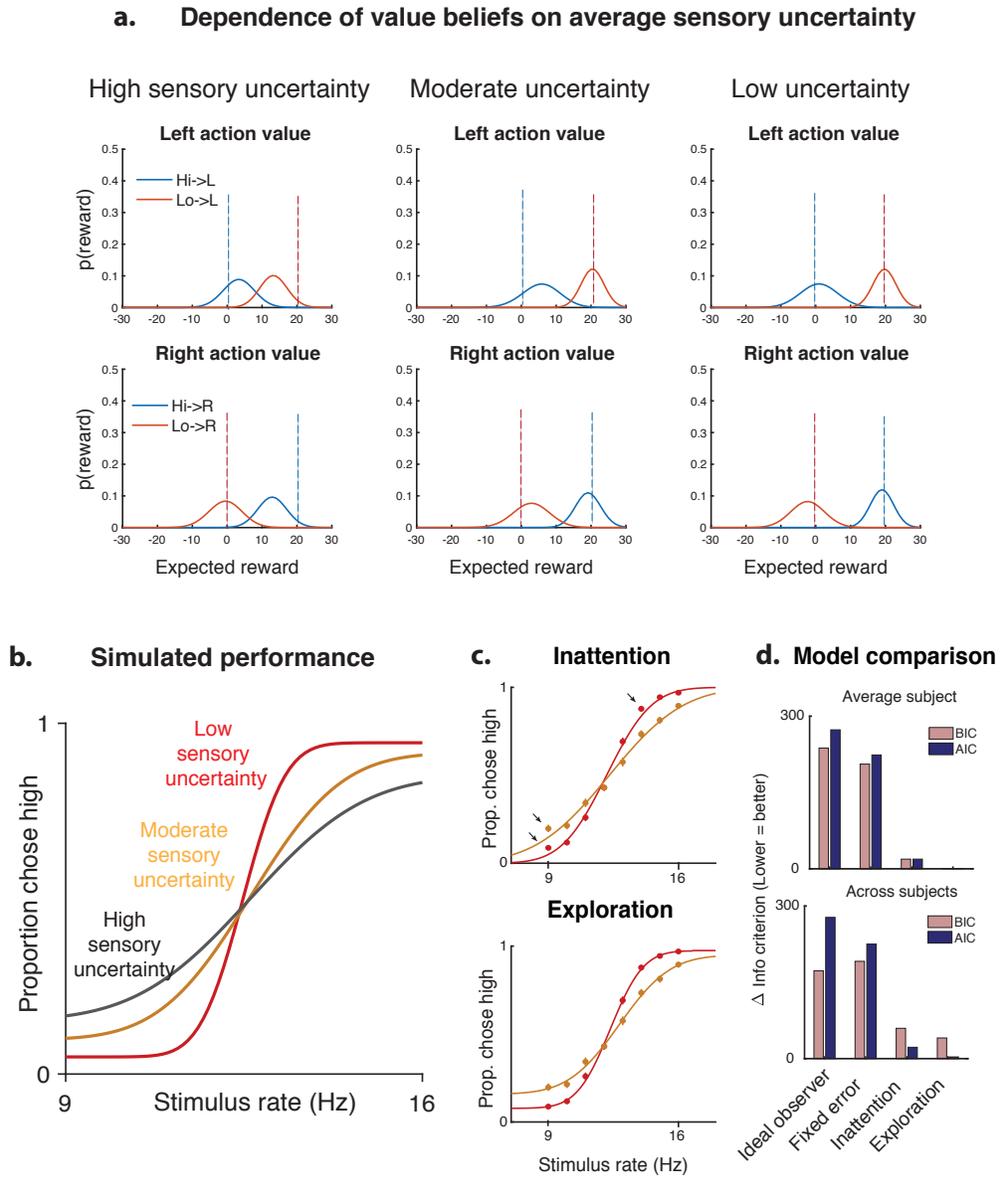
Learning the expected rewards is further complicated by uncertainty in the stimulus category - if the animal receives a reward for rightward actions on a trial on which the stimulus was extremely ambiguous, then it will not know whether to attribute the reward to Rightward actions in a High state or Rightward actions in a low state (In fact, the Bayesian update rule hedges its bets and attributes it slightly to both, updating both expected rewards). This problem will be more severe in conditions with higher perceptual uncertainty, leading to highly overlapping beliefs about expected rewards, and consequently more exploration on these conditions. This predicts that conditions with higher perceptual uncertainty (e.g. unisensory or neutral trials) should encourage more exploration, giving rise to more frequent lapses (Figure 4.2).

### 4.1.3 Model comparison

As a result, the uncertainty-guided exploration model predicts an increase not only in the  $\sigma$  but also in lapses on neutral trials, just as we observed- in fact it predicts that both these parameters should match those on auditory trials. This model fit the data well. Model comparison using BIC and AIC both favored the exploration model over the inattention model for average data as well as across individual subjects

## 4.2 Testing the predictions of the exploration model

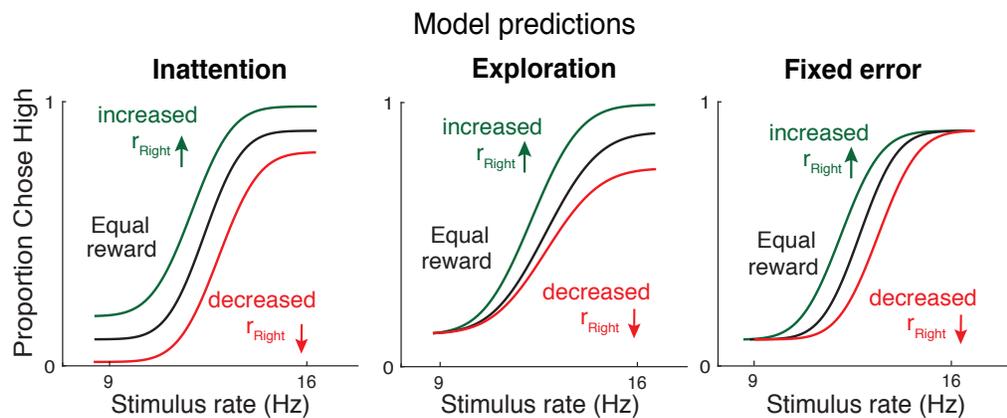
One of the key features of the uncertainty-guided exploration model is that lapses are exploratory choices made with full knowledge of the stimulus, and should depend only on the expected rewards associated with that stimulus category. This is in stark contrast to the inattention model and many other kinds of disengagement, in which lapses are caused by



**FIGURE 4.2: Perceptual uncertainty increases exploration.** (a) Beliefs of a Bayesian learner with Thompson sampling, under high (left), medium (center) or low (right) perceptual uncertainty. (c) Decreased uncertainty allows better disambiguation between low or high, leading to more separable beliefs, less exploration and lower lapses.

the observer disregarding the stimulus, and hence lapses at the two extreme stimulus levels are both influenced by a common underlying guessing process that depends on expected rewards from both stimulus categories. This is also in contrast with fixed motor error or  $\epsilon$ -greedy models in which lapses are independent of expected reward.

Therefore, a unique prediction of the exploration model is that selectively manipulating expected rewards associated with one of the stimulus categories should only affect lapses at one extreme of the psychometric function, whereas inattention and other kinds of disengagement predict that both lapses should be affected, and fixed error models predict that neither should be affected (Figure 4.3).



#### 4.2.1 Manipulating reward magnitude

To experimentally test these predictions, we tested rats on the rate discrimination task with asymmetric rewards. Instead of rewarding high and low rate choices equally, we increased the water amount on the reward port associated with high-rates (rightward choices) so it was 1.5 times larger than before, without changing the reward on the the low-rate side (leftward choices). In a second rat cohort we did the opposite: we devalued the choices associated

with high-rate trials by decreasing the water amount on that side port so it was 1.5 times smaller than before, without changing the reward on the low-rate side. The animals' behavior on the asymmetric-reward task matched the predictions of the exploration model (Figure 4.4). Increasing the reward size on choices associated with high-rates led to a decrease in lapses for the highest rates and no changes in lapses for the lower rates (n=3 rats, 6976 trials). Decreasing the reward of choices associated with high-rates led to an increase in lapses for the highest rates and no changes in lapses for the lower rates (n=3 rats, 11164 trials). This shows that both increasing and decreasing the value of one of the actions has an asymmetric effect on lapse probabilities that does not match the inattention model.

#### 4.2.2 Manipulating reward probability

As an additional test of the model, we manipulated expected rewards by probabilistically rewarding incorrect i.e. leftward choices on high rate trials with a probability of 0.5, while leaving all other rewards unchanged (Figure 4.4). The exploration model predicts that this should selectively increase the value of leftward actions on high rate trials, increasing lapses on high rates. Indeed, this is what we observed (n=5 animals, 347537 trials), and the effect was strikingly similar to the decreased reward experiment, even though the two manipulations affect high rate action values through changes on opposite reward ports. Moreover, this suggests that lapses reflect changes in action value caused by changing either reward magnitudes or reward probabilities, as one would expect from the exploration model.

#### 4.2.3 Effects of uncertainty

To confirm that the asymmetric changes in lapse rate that we observed were truly driven by uncertainty, we examined performance on randomly interleaved "sure bet" trials on which the uncertainty was very low (Figure 4.4). On these trials, a pure tone was played during

the fixation period, after which an LED at one of the side ports was clearly illuminated, indicating a reward. Sure-bet trials comprised 6% of the total trials, and as with the rate discrimination trials, left and right trials were interleaved. Owing to the low uncertainty, the model predicts that very little exploration would be required in this condition, and that animals would very quickly reach perfect performance on these trials. Importantly, our model predicts that performance on “sure-bet” trials would be unaffected by imbalances in reward magnitude.

In keeping with this prediction, on sure-bet trials, performance was near perfect (rightward probabilities of 0.003 [0.001,0.01] and 0.989 [0.978,0.995] on go-left and go-right trials respectively), and unaffected following reward manipulations (Rightward probabilities of 0.004 [0.001, 0.014] and 0.996 [0.986,0.999] on increased reward, 0.006 [0.003,0.012] and 0.99 [0.983,0.994] on decreased reward). This suggests that the effect of value on lapses is restricted to uncertain situations that encourage subjects to explore, rather than exploit. Further, because sure-bet trials were interleaved with more uncertain trials, their near-perfect performance indicates that uncertainty can be estimated on the timescale of individual trials.

#### 4.2.4 Effects of reward history

The subjective value of actions may naturally change with experience, even without the explicit reward manipulations described above. Throughout training, the animal uses outcomes of previous trials to learn and update the expected rewards from various actions, allowing it to learn the rules of the task (i.e. that high rate trials are rewarded on the right and so on). If such learning processes continue to persist in trained animals, either due to incomplete training (Law and Gold, 2009), uncertainty in feedback, forgetting over time (Gershman, 2015; Drugowitsch and Pouget, 2018), or perceived volatility (Yu and Cohen, 2009), then the outcomes of previous trials should continue to affect subsequent trials even

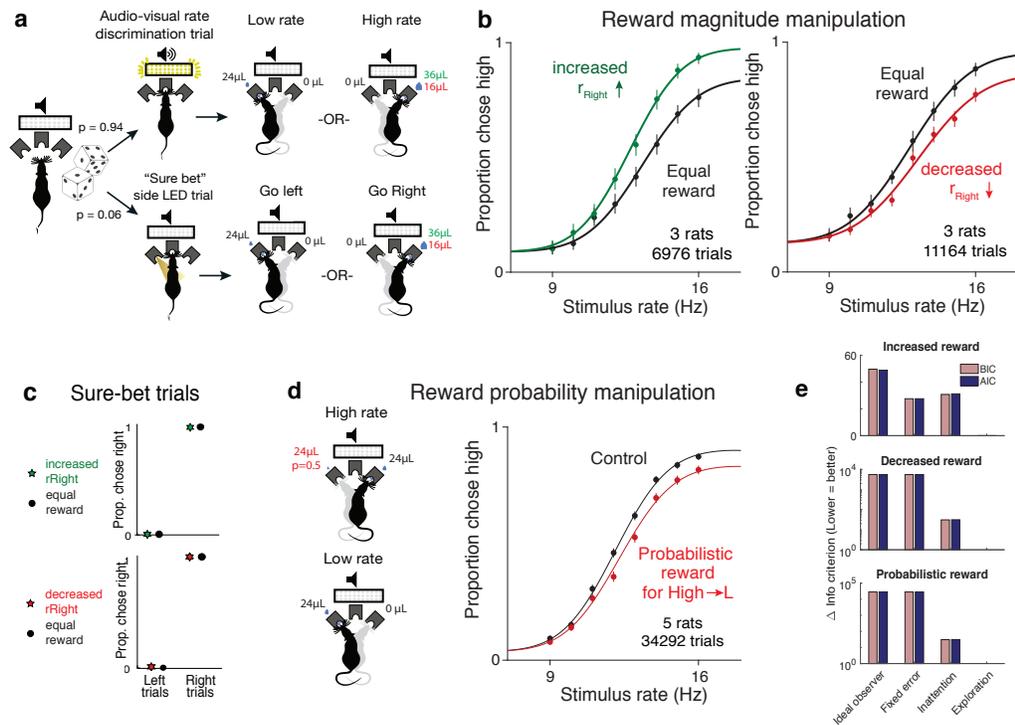


FIGURE 4.4: **Reward manipulations confirm exploration model predictions.** (a) Schematic of rate discrimination trials (top) and sure bet trials (bottom) with increased or decreased reward magnitude. (b) Behavior of 3 rats for increased (left) or decreased (right) reward, with exploration model fits (solid lines). (c) No effect on sure-bet trials. (d) Reward probability manipulation (left) and behavior of 5 rats (right) with exploration model fits (solid lines). (e) Model comparison showing exploration model best fits all manipulations

in trained animals, as has been observed in a number of studies (Busse et al., 2011; Lak et al., 2018; Mendonca et al., 2018; Odoemene et al., 2018; Pinto et al., 2018; Scott et al., 2015). The action value of rightward choices should increase following a rightward success and decrease following a rightward failure, predicting the same asymmetric changes in lapses as reward magnitude manipulations. As predicted, trials following rewarded and unrewarded rightward choices showed decreased and increased high-rate lapses, respectively (Figure 4.5).

#### 4.2.5 Effects of contingency reversal

The modulation of lapses with reward history suggests that they reflect a dynamic process of action value updating. However, such continual updating does not confer any advantage in our task, since the contingencies are actually static. So to test the predictions of the model in a situation where continued learning and exploration is adaptive, we reversed the reward contingency in a separate cohort of rats, such that high rates were now rewarded on the left instead of the right. The model predicted that as the uncertainty in action values increases due to the reversal and then decreases with re-learning, the proportion of exploratory choices should increase and then decrease. Indeed, this is what we observed (Figure 4.5). Taken together, manipulations of value match the predictions of the uncertainty-dependent exploration model.

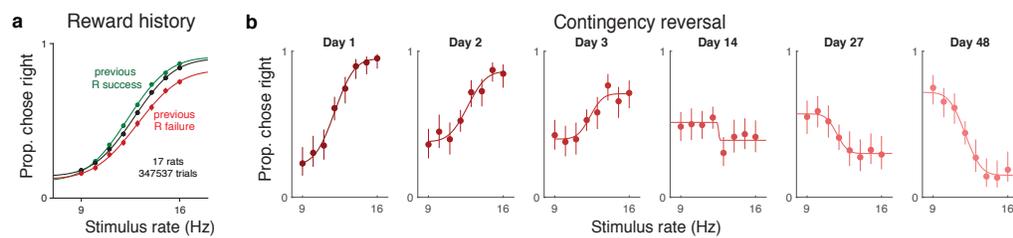


FIGURE 4.5: **Effects of learning on lapses.** Changing action values due to (a) previous outcomes and (b) contingency reversal change lapses as expected from exploration model.

### 4.3 Discussion

Balancing exploration and exploitation is computationally challenging, and uncertainty-dependent exploration or Thompson sampling is a well-known, elegant heuristic for achieving this balance. It has been shown to be asymptotically optimal in partially observable environments (Leike et al., 2016) and can be implemented through a sampling scheme where the subject samples action values from a learnt distribution and then maximizes with respect to the sample (Gershman, 2018).

This strategy predicts that conditions with higher uncertainty should have higher exploration, and consequently higher lapse rates, explaining the pattern of lapse rates we observed on unisensory vs. multisensory trials as well as on neutral vs. matched trials. A strong prediction of uncertainty guided exploration is that the animal should always exploit on conditions with no uncertainty, as we observed on sure-bet trials.

A unique prediction of the exploration model is that it predicts lapse rates will sometimes change asymmetrically for left and right decisions. For instance, changing the value associated with one of the decisions (eg. high rate) should only affect lapses associated with that decision - predicting fewer lapses on high rates if the rightward reward is increased, and more lapses if it is decreased, or if leftward decisions are probabilistically rewarded on high rates. These predictions are borne out, and rightward successes or failures on the previous trial have a similar effect.

The model also predicts that exploration, and consequently lapses, should decrease with training as the animal becomes more certain of the rules and expected rewards, explaining the changes we see after reversal learning, as well as training-dependent effects reported in primates (Law and Gold, 2009).

## Chapter 5

# Leveraging lapses to infer the stage of decision-making deficits

### 5.1 Informativeness of lapses in perturbation experiments

The results of the reward experiments suggest that disrupting areas that confer value to actions should asymmetrically bias lapses, in contrast to disruptions of areas that encode sensory evidence, which should lead to horizontal biases without affecting lapses, or motor disruptions that simply make one of the actions harder to perform, which should affect both lapses. Crucially, in the absence of lapses, all three of these disruptions would look identical, producing a horizontal shift (Figure 5.1). This suggests that lapses could actually be informative about the stage of involvement of brain regions.

### 5.2 Identifying candidate multisensory areas

Two candidate areas that we sought out to test in our multisensory task were secondary motor cortex (M2) and posterior striatum (pStr), both of which receive convergent input from primary visual and auditory cortices (Figure 5.2, results of simultaneous anterograde tracing

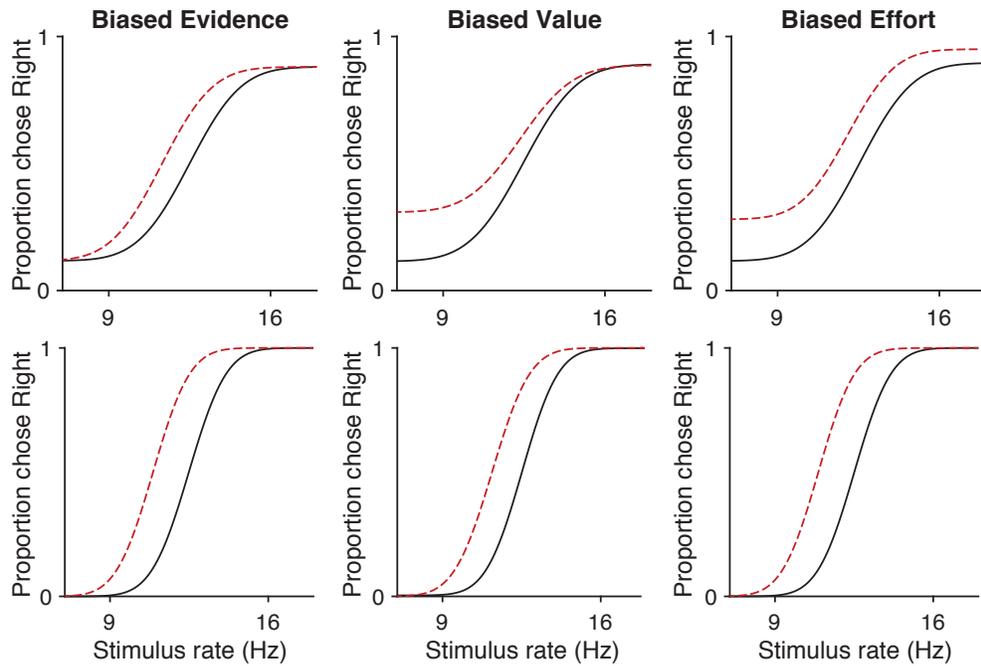


FIGURE 5.1: **Lapses are informative about decision deficits.** Biasing decisions at various stages predicts different effects on lapses - biased evidence (top left) predicts no change in lapses, biased value (top center) predicts a change in one of the lapses and biased effort to move (top right) predicts both lapses will change. However in the absence of lapses (bottom), all three manipulations have the same effect.

from V1 and A1; experiments performed in collaboration with Dr. Lital Chartarifsky. Also see Jiang and Kim, 2018; Barthas and Kwan, 2017). Previous studies have shown effects on lapses following inactivation of both these areas in auditory tasks in rats (Erlich et al., 2015; Guo et al., 2018). These were interpreted as effects arising from either leaky accumulation (Erlich et al., 2015), post-categorization biases (Piet et al., 2017) or perceptual biases (Guo et al., 2018). These effects were very similar to the effects of manipulating reward in our task, hinting that these effects may actually arise from biased action values. However, since these deficits were only demonstrated during auditory decision-making, these studies did not afford the opportunity to distinguish sensory modality-specific deficits from those that should generalize across modalities (e.g., visual, multisensory, auditory) like value deficits.

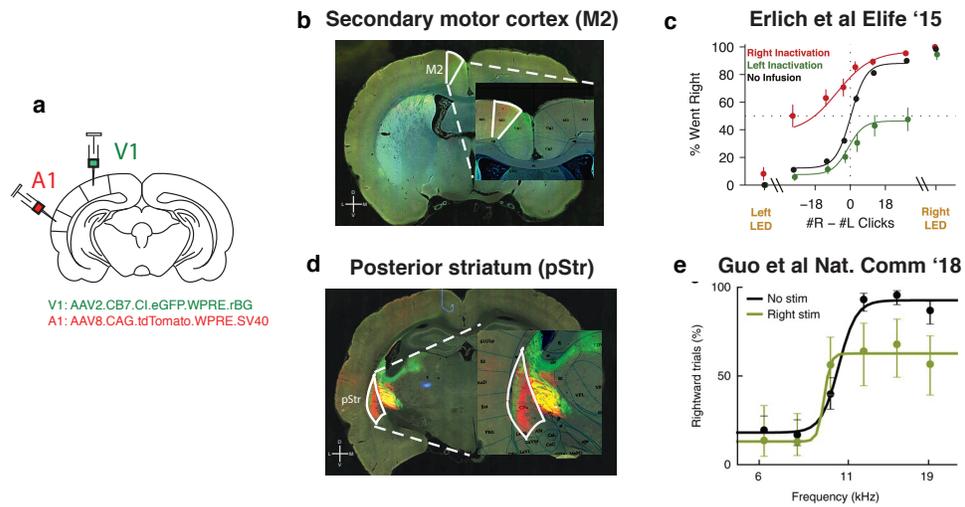
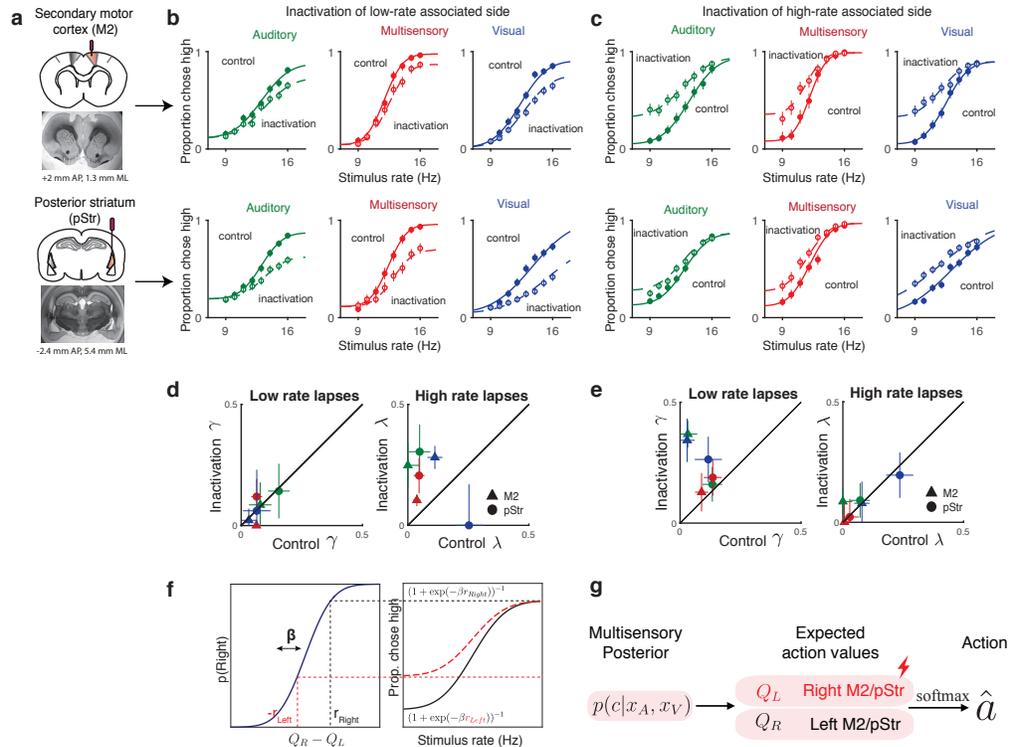


FIGURE 5.2: **Identifying candidate multisensory areas.** (a) Schematic outlining experiment to trace projections from primary visual and auditory cortices. (b) Convergent projections in secondary motor cortex (M2) (c) Effect of M2 inactivations on auditory decisions, reproduced from Erlich et al '15. (d) Convergent projections in posterior striatum (pStr) (d) Effect of pStr on auditory decisions, reproduced from Guo et al '18

### 5.3 Prefrontal, striatal inactivations resemble value deficit

To test whether pStr and M2 have a modality-independent role in perceptual decisions, we suppressed activity of neurons in each of these areas using muscimol, a  $GABA_A$  agonist, during our multisensory rate discrimination task. We implanted bilateral cannulae in M2 (Supplementary Fig. 6b;  $n = 5$  rats; +2 mm AP 1.3 mm ML, 0.3 mm DV) and pStr (Supplementary Fig. 6a;  $n = 6$  rats; -3.2 mm AP, 5.4 mm ML, 4.1 mm DV) (Fig. 5a). On control days, rats were infused unilaterally with saline, followed by unilateral muscimol infusion the next day (M2: 0.1-0.5  $\mu\text{g}$ , pStr 0.075-0.125  $\mu\text{g}$ ). We compared performance on the multisensory rate discrimination task for muscimol days with preceding saline days (Surgeries and infusions performed by Dr. Lital Chartarifsky).



**FIGURE 5.3: Prefrontal, striatal inactivations resemble value deficit.** (a) Schematic and histology of M2 (top) and pStr (bottom) cannulae implants. (b,c) Effect of inactivating the side associated with high rates (b) or low rates (c) on auditory (green), multisensory (red) or visual (blue) decisions. Solid lines are exploration model fits with contralateral devaluation. (d,e) Inactivation on low rate (d) or high-rate (e) side only affects contralateral lapses. (f) Schematic of exploration model predictions from contralateral devaluation. (g) Proposed role of M2, pStr in encoding lateralized action values based on multisensory beliefs.

### 5.3.1 Asymmetric effects on lapses

Inactivation of the side associated with low-rate choices biased the animals to make more low-rate choices (Figure 5.3; left 6 panels: empty circles, inactivation sessions; full circles, control sessions) and inactivation of the side associated with high-rates biased them to make more high-rate choices (Figure 5.3; right 6 panels). The inactivations largely affected lapses on the contralateral side, while sparing those on the ipsilateral side. These results recapitulated previous findings, and were strikingly similar to the effects we observed following reward manipulations (as seen in Fig. 4.4). These effects were seen across areas and modalities, suggesting that pStr and M2 are part of a modality-independent circuit for decision-making.

### 5.3.2 Model comparison

Fitting the data with the exploration model revealed that the effects on lapses could be captured by scaling the contralateral action value by a single parameter across modalities (Figure 5.3, Joint fits to control (solid lines) and inactivation trials (dotted lines) across modalities, differing only by a single parameter), similar to the reward manipulation experiments. Animals that were inactivated on the side associated with high rates showed increased low rate lapses, but lapses did not change for high rates. This was consistent across areas and modalities. Animals that were inactivated on the side associated with low rates showed the opposite effect: increased lapses on high rate trials and no change in lapses for low rate trials.

### 5.3.3 Sensory and motor controls

To confirm that this effect was independent of the associated stimulus, some rats were trained on a reverse contingency regimen in which high rates were rewarded on the left

side. The effects were consistent across both groups and always resembled a devaluation of contralateral actions (Figure 5.4);.

To determine whether changes in decision-making simply reflected motor impairments that drove a tendency to favor the ipsilateral side, we compared behavior on the sure-bet task described previously. Performance was spared on these trials: rats made correct rightward and leftward choices regardless of the side that was inactivated (Figure 5.4). This suggests that behavioral effects were restricted to situations in which there was uncertainty about the correct outcome. We also looked at multiple movement parameters such as wait time in the center port and movement times to ipsilateral and contralateral reward ports. There were no significant effects on movement parameters, suggesting that effects on decision outcome were not due to a muscimol-induced motor impairment.

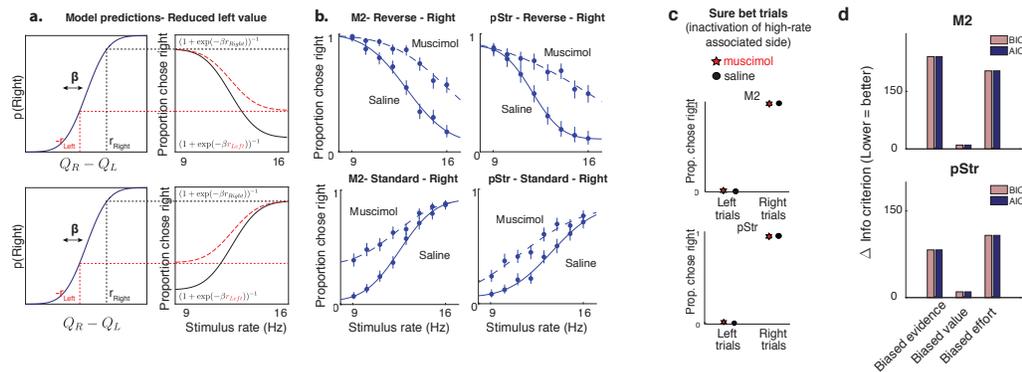


FIGURE 5.4: **Inactivation effects are not sensory or motor.** (a) Predictions of devaluation of leftward actions on decisions with standard (bottom) and reversed (top) stimulus-response contingencies. (b) Data from right M2 (left) or right pStr (right) inactivations on standard (bottom) or reversed (top) contingencies. (c) No effect on sure-bet trials. (d) Model comparison shows that biased contralateral value fits better than biased contralateral evidence or effort.

Finally, a model comparison revealed that a fixed contralateral value deficit captured the inactivation effects much better than a fixed reduction in contralateral sensory

evidence or a fixed increase in contralateral motor effort, both for M2 and pStr. In uncertain conditions, this reduced contralateral value gives rise to more exploratory choices and hence more lapses on one side, but doesn't affect the other side, or sure-bet trials on which the animals largely exploit.

#### 5.3.4 Discussion

The exploration model predicts that perturbing the decision-making process at different stages should have differential effects on lapses. Using this knowledge, we were able to conclude that the asymmetric effects of prefrontal, striatal inactivations reflect a deficit in contralateral action value, suggesting that M2 and pStr have a lateralized, modality-independent role in computing the expected value of actions based on incoming multisensory information.

This interpretation can provide a unified explanation for the asymmetric effects on lapses reported in a number of tasks in rats and mice (Erlich et al., 2015; Wang et al., 2018; Guo et al., 2018), which have been variously interpreted as post-categorization biases, response biases and evidence biases. It can also account for the puzzling observation that on a task where left and right actions are not coupled i.e. where animals have a third, no-go option available, unilateral inactivation of M2 reduces contralateral responding without increasing ipsilateral responding (Zatka-Haas et al., 2019). The shape of the deficit also suggests that the animals' evidence processing is intact and the deficit arises from devaluation of actions, explaining why temporally specific inactivation of M2 only disrupts decisions towards the end of evidence accumulation (Hanks et al., 2015). Finally, this interpretation reconciles the deficits on these perceptual tasks with the established roles of these areas in encoding action value (Barthas and Kwan, 2017; Lee et al., 2015) during value-based decisions.

## Chapter 6

# Conclusion and Perspectives

### 6.1 Summary of findings

Perceptual decision-makers have long been known to display a small fraction of errors even on easy trials. Until now, these “lapses” were largely regarded as a nuisance and lacked a comprehensive, normative explanation. In this thesis, I proposed a novel explanation for lapses: that they reflect a strategic balance between exploiting known rewarding options and exploring uncertain ones. Our model makes strong predictions for lapses under diverse decision-making contexts, which I have tested here. First, the model predicts more lapses on conditions with higher uncertainty, such as unisensory (Chapter 2) or neutral (Chapter 3), compared to multisensory or sure-bet conditions. Second, the model predicts that asymmetric reward manipulations should only affect lapses on one side, sparing decisions to the other side and sure-bet trials (Chapter 4). Finally, the model predicts that lapses should be affected by perturbations to brain regions that encode action value. Accordingly, we observed that unilateral inactivations of secondary motor cortex and posterior striatum similarly affected lapses on one side across auditory, visual and multisensory trials (Chapter 5). Taken together, our model and experimental data argue strongly that far from being a nuisance, lapses are informative about animals’ subjective action values and reflect a trade-off

between exploration and exploitation.

## 6.2 The influence of value on perceptual decisions

Considerations of value have provided many useful insights into aspects of behavior that seem sub-optimal at first glance from the perspective of perceptual ideal observers. For instance, many perceptual tasks are designed with accuracy in mind - defining an ideal observer as one who maximizes accuracy, in line with classical signal detection theory. However, in practice, the success or failure of different actions may be of unequal value to subjects, especially if reward or punishment is delivered explicitly, as is often the case with non-human subjects. This may give rise to biases that can only be explained by an observer that maximizes expected utility (Dayan and Daw, 2008). Similarly, reward outcomes on a given trial can influence decisions about stimuli on subsequent trials through reinforcement learning, giving rise to serial biases. These biases occur even though the ideal observer should treat the evidence on successive trials as independent (Busse et al., 2011; Lak et al., 2018). Finally, when subjects can control how long they sample the stimulus, subjects maximizing reward rate may choose to make premature decisions, sacrificing accuracy for speed (Bogacz et al., 2006).

In this thesis, I take further inspiration from considerations of value to provide a novel account for lapses. I leveraged a well known phenomenon in value-based decisions: uncertainty dependent exploration. Until now, this phenomenon has not been considered a candidate explanation for lapses in perceptual decisions. Our results argue that lapses are not simply accidental errors made as a consequence of attentional “blinks” or motor “slips”, but can reflect a deliberate, internal source of behavioral variability that facilitates learning and information gathering under uncertain or non-stationary environments.

### 6.3 Insights into individual differences

The exploration model of lapses suggests that variability seen in lapses across conditions within subjects, and across healthy subjects and patients (Bertolini et al., 2015; Mihali et al., 2018; Nikbakht et al., 2018) might reflect condition- and subject- specific modulations of the exploration-exploitation tradeoff, offering an opportunity to quantify individual differences. It can also potentially explain why children have higher lapse rates - they have been shown to be more exploratory in their decisions than adults (Lucas et al., 2014). The model also predicts that within an individual subject and condition, exploration, and consequently lapses, should decrease with training as the animal becomes more certain of the rules and expected rewards, explaining training-dependent effects reported in primates (Law and Gold, 2009).

### 6.4 Insights into the role of prefrontal cortex, striatum

A unique prediction of the exploration model is that it predicts lapse rates can change asymmetrically if the action value for one of the decisions is selectively affected. This suggests that the asymmetric effects on lapses seen during unilateral inactivations of prefrontal and striatal regions arises from a selective devaluation of contralateral actions. This interpretation reconciles a number of studies that have found asymmetric effects of inactivating these areas during perceptual decisions (Erlich et al., 2015; Zatka-Haas et al., 2019; Wang et al., 2018; Guo et al., 2018) with their established roles in encoding action value (Barthas and Kwan, 2017; Lee et al., 2015) during value-based decisions.

## 6.5 Uncertainty-guided exploration and dopamine

An open question that remains is how the brain might tune the degree of exploration in proportion to uncertainty. An intriguing candidate for this is dopamine, whose tonic levels have been shown to modulate exploration in mice on a lever-press task (Beeler et al., 2010), and context-dependent song variability in songbirds (Leblois, Wendel, and Perkel, 2010). Dopaminergic genes have been shown to predict individual differences in uncertainty-guided exploration in humans (Frank et al., 2009), and dopaminergic disorders such as Parkinson's disease have been shown to disrupt the uncertainty-dependence of lapses across conditions (Bertolini et al., 2015). Patients with ADHD, another disorder associated with dopaminergic dysfunction, have been shown to display both increased perceptual variability and increased task-irrelevant motor output, a measure that correlates with lapses (Mihali et al., 2018). A promising avenue for future studies is to leverage the informativeness of lapses and the precise control of uncertainty afforded by multisensory tasks, in conjunction with perturbations or recordings of dopaminergic circuitry, to further elucidate the connections between perceptual and value-based decision making systems.

## Appendix A

# Methods

### A.1 Behavioral training

*Animal Subjects and Housing* All animal procedures and experiments were in accordance with the National Institutes of Health's Guide for the Care and Use of Laboratory Animals and were approved by the Cold Spring Harbor Laboratory Animal Care and Use Committee. Experiments were conducted with 34 adult male and female Long Evans rats (250-350g, Taconic Farms) that were housed with free access to food and restricted access to water starting from the onset of behavioral training. Rats were housed on a reversed light-dark cycle; experiments were run during the dark part of the cycle. Rats were pair-housed during the whole training period.

*Animal training and behavioral task* Rats were trained following previously established methods (Raposo 2012, Sheppard 2013, Raposo 2014, Licata 2017). Briefly, rats were trained to wait in the center port for 1000 ms while stimuli were presented, and to associate stimuli with left/right reward ports. Stimuli for each trial consisted of a series of events: auditory clicks from a centrally positioned speaker, full-field visual flashes, or both together. Stimulus events were separated by either long (100 ms) or short (50 ms) intervals. For the easiest trials, all inter-event intervals were identical, generating rates that were 9 events/s (all

long intervals) or 16 events/s (all short intervals). More difficult trials included a mixture of long and short intervals, generating stimulus rates that were intermediate between the two extremes and therefore more difficult for the animal to judge. The stimulus began after a variable delay following when the rat's snout broke the infrared beam in the center port. The length of this delay was selected from a truncated exponential distribution ( $\lambda = 30$  ms, minimum = 10 ms, maximum = 200 ms) to generate an approximately flat hazard function. The total time of the stimulus was usually 1000 ms. Trials of all modalities and stimulus strengths were interleaved. For multisensory trials, the same number of auditory and visual events were presented (except for a subset of 'neutral' trials). Auditory and visual stimulus event times were generated independently, as our previous work has demonstrated that rats make nearly identical decisions regardless of whether stimulus events are presented synchronously or independently (Raposo et al., 2012). For most experiments, rats were rewarded with a drop of water for moving to the left reward port following low-rate trials and to the right reward port following high rate trials. For muscimol inactivation experiments half of the rats were rewarded according to the reverse contingency. Animals typically completed between 700 and 1,200 trials per day. Most experiments had 18 conditions (3 modalities x 8 stimulus strengths), leading to 29-50 trials per condition per day.

## **A.2 Behavioral manipulations**

### **A.2.1 Congruence manipulation**

To probe the effect of perceptual uncertainty on lapses, rats received "catch" trials consisting of multisensory neutral trials, where only the auditory modality provided evidence for a particular choice, whereas the visual modality provided evidence that was so close to the category boundary (12 Hz) that it did not support one choice or the other (Raposo et al., 2012)

### A.2.2 Reward manipulation

To probe the effect of value on lapses, we manipulated either reward magnitude or reward probability associated with high rates, while keeping low rate trials unchanged. To increase or decrease reward magnitude associated with high rates, the amount of water dispensed on the right port was increased or decreased to 36ul or 16 ul respectively, while the reward on the left port was maintained at 24 ul. To manipulate reward probability, we occasionally rewarded rats on the (incorrect) left port on high rate trials with a probability of 0.5. The right port was still rewarded with a probability of 1 on high rates, and reward probabilities on low rate trials were unchanged (1 on the left port, 0 on the right).

### A.3 Tracing experiments

*Viral injections*- 2 rats, 15 weeks of age, were anesthetized and placed in a stereotaxic apparatus (Kopf Instruments). Small craniotomies were made in the center of primary visual cortex (V1; 6.9mm posterior to Bregma, 4.2mm to the right of midline) and primary auditory cortex (A1; 4.7mm posterior to Bregma, 7mm to the right of midline). Small durotomies were performed at each craniotomy and virus was pressure injected at depths of 600, 800, and 1000  $\mu\text{m}$  below the pia (150 nL/depth). Virus injections were performed using Drummond Nanoject III, which enables automated delivery of small volumes of virus. To minimize virus spread, the Nanoject was programmed to inject slowly: fifteen 10 nL boluses, 30 seconds apart. Each bolus was delivered at 10 nL/sec. 2-3 minutes were allowed following injection at each depth to allow for diffusion of virus. The AAV2.CB7.CI.EGFP.WPRE.RBG construct was injected in V1, and the AAV2.CAG.tdTomato.WPRE.SV40 construct was injected in A1. Viruses were obtained from the University of Pennsylvania vector core.

## A.4 Neural manipulations

All rats subject to surgery were anesthetized with 1%-3% isoflurane. Isoflurane anesthesia was maintained by monitoring respiration, heart rate, oxygen and CO<sub>2</sub> levels, as well as foot pinch responses throughout the surgical procedure. Ophthalmic ointment was applied to keep the eyes moistened throughout surgery. After scalp shaving, the skin was cleaned with 70% ethanol and 5% betadine solution. Lidocaine solution was injected below the scalp to provide local analgesia prior to performing scalp incisions. Meloxicam (5mg/ml) was administered subcutaneously (2mg/kg) for analgesia at the beginning of the surgery, and daily 2-3 days post-surgery. The animals were allowed at least 7 days to recover before behavioral training.

***Cannulae implants*** Rats were anesthetized and placed in the stereotax as described above. After incision and skull cleaning, 2 skull screws were implanted to add more surface area for the dental cement. For striatal implants, two craniotomies were made, one each side of the skull (3.2mm posterior to Bregma; 5.4mm to the right and left of midline). Durotomies were performed and a guide cannula (22 gauge, 8.5 mm long; PlasticsOne) was placed in the brain, 4.1mm below the pia at each craniotomy. For secondary motor cortex implants, one large craniotomy spanning the right and left M2 was performed (~5mm x ~2mm in size centered around 2mm anterior to Bregma and 3.1mm to the right and left of midline). A durotomy was performed and a double guide cannula (22 gauge, 4mm long; PlasticsOne) was placed in the brain, 300um below the pia. The exposed brain was covered with sterile Vaseline and cannulae were anchored to the skull with dental acrylic (Relyx). Single or double dummy cannulae protruding 0.7 mm below the guide cannulae were inserted.

#### **A.4.1 Inactivation with muscimol**

Rats were lightly anesthetized with isoflurane. Muscimol was unilaterally infused into pStr or M2 with a final concentration of 0.075-0.125  $\mu\text{g}$  and 0.1-0.5  $\mu\text{g}$ , respectively. A single/double-internal cannula (PlasticsOne), connected to a 2  $\mu\text{l}$  syringe (Hamilton microliter syringe, 7000 series), was inserted into each previously implanted guide cannula. Internal cannulae protruded 0.5mm below the guide. Muscimol was delivered using an infusion pump (Harvard PHD 22/2000) at a rate of 0.1  $\mu\text{l}/\text{minute}$ . Internal cannulae were kept in the brain for 3 additional minutes to allow for diffusion of muscimol. Rats were removed from anesthesia and returned to cages for 15 minutes before beginning behavioral sessions. The same procedure was used in control sessions, where muscimol was replaced with sterile saline.

#### **A.4.2 Histology**

At the conclusion of inactivation experiments, animals were deeply anesthetized with Euthasol (pentobarbital and phenytoin). Animals were perfused transcardially with 4% paraformaldehyde. Brains were extracted and post-fixed in 4% paraformaldehyde for 24-48 hours. After post-fixing, 50-100  $\mu\text{m}$  coronal sections were cut on a vibratome (Leica) and imaged.

## A.5 Behavioral analysis

### A.5.1 Psychometric curves

Descriptive four-parameter psychometric functions were fit to choice data using the Palamedes toolbox (Prins and Kingdom, 2018). Psychometric functions were parameterized as:

$$\psi(x; \mu, \sigma, \gamma, \lambda) = \phi(x; \mu, \sigma)(1 - \lambda - \gamma) + \gamma \quad (\text{A.1})$$

where  $\gamma$  and  $\lambda$  are the lower and upper asymptote of the psychometric function, which parameterize the lapse rates on the left and to the right, respectively;  $\phi$  is a cumulative normal function;  $x$  is the event rate, i.e. the number of flashes or beeps presented during the one second stimulus period;  $\mu$  parameterizes the  $x$ -value at the midpoint of the psychometric function and  $\sigma$  describes the inverse slope. 95% Confidence intervals on these parameters were generated via bootstrapping based on 1000 simulations.

## A.6 Modeling

### A.6.1 Ideal observer model

We can specify an ideal observer model for our task using Bayesian Decision Theory (Dayan and Daw, 2008). This observer maintains probability distributions over previously experienced stimuli and choices, computes the posterior probability of each action being correct given its observations and picks the action that yields the highest expected reward.

Let the true category on any given trial be  $c_{true}$ , the true stimulus rate be  $s_{true}$  and the animal's noisy visual and auditory observations of  $s_{true}$  be  $x_V$  and  $x_A$ , respectively. We assume that the two sensory channels are corrupted by independent gaussian noise with

standard deviation  $\sigma_A$  and  $\sigma_V$ , respectively, giving rise to conditionally independent observations.

$$\begin{aligned} p(x_A|s_{true}) &= \mathcal{N}(s_{true}, \sigma_A), & p(x_V|s_{true}) &= \mathcal{N}(s_{true}, \sigma_V), \\ p(x_A, x_V|s_{true}) &= p(x_A|s_{true})p(x_V|s_{true}) \end{aligned} \quad (\text{A.2})$$

The ideal observer can use this knowledge to compute the likelihood of seeing the current trial's observations as a function of the hypothesized stimulus rate  $s$ . This likelihood  $\mathcal{L}$  is a gaussian function of  $s$  with a mean given by a weighted sum of the observations  $x_A$  and  $x_V$ :

$$\begin{aligned} \mathcal{L}(s) &= p(x_A, x_V|s) = p(x_A|s)p(x_V|s) \\ &\propto \mathcal{N}(\mu_M, \sigma_M) \\ \mu_M &= w_A x_A + w_V x_V \\ \sigma_M &= (\sigma_A^{-2} + \sigma_V^{-2})^{-\frac{1}{2}} \\ w_A &= \frac{\sigma_M^2}{\sigma_A^2}, & w_V &= \frac{\sigma_M^2}{\sigma_V^2} \end{aligned} \quad (\text{A.3})$$

The likelihood of seeing the observations as a function of the hypothesized category  $c$ , is given by marginalizing over all possible hypothesized stimulus rates. Let the experimentally imposed category boundary be  $\mu_0$ , such that stimulus rates are considered high when  $s > \mu_0$  and low when  $s < \mu_0$ . Then,

$$\begin{aligned} \mathcal{L}(c = \text{High}) &= p(x_A, x_V|c = \text{High}) \\ &= \int_s p(x_A, x_V, s|c = \text{High}) ds \\ &= \int_s p(x_A, x_V|s)p(s|c = \text{High}) ds \quad \because x_A, x_V \perp c|s \\ &= \int_{s > \mu_0} p(x_A, x_V|s) ds \\ &\propto 1 - \Phi(\mu_0; \mu_M, \sigma_M) \end{aligned} \quad (\text{A.4})$$

where  $\Phi$  is the cumulative normal function. Using Bayes' rule, the ideal observer can then compute the probability that the current trial was high or low rate given the observations, i.e. the posterior probability.

$$\begin{aligned}
 p(c|x_A, x_V) &= \frac{p(x_A, x_V|c)p(c)}{p(x_A, x_V)} \\
 \implies p(c = \text{High}|x_A, x_V) &\propto p_{\text{High}}(1 - \Phi(\mu_0; \mu_M, \sigma_M)) \\
 \implies p(c = \text{Low}|x_A, x_V) &\propto p_{\text{Low}}\Phi(\mu_0; \mu_M, \sigma_M)
 \end{aligned} \tag{A.5}$$

where  $p_{\text{High}}$  and  $p_{\text{Low}}$  are the prior probabilities of high and low rates respectively. The expected value  $Q(a)$  of choosing right or left actions (also known as the action values) is obtained by marginalizing the learnt value of state-action pairs  $q(c, a)$  over the unobserved state  $c$ .

$$\begin{aligned}
 Q(a = R) &= p(\text{High}|x_A, x_V)q(\text{High}, R) + p(\text{Low}|x_A, x_V)q(\text{Low}, R) \\
 Q(a = L) &= p(\text{High}|x_A, x_V)q(\text{High}, L) + p(\text{Low}|x_A, x_V)q(\text{Low}, L)
 \end{aligned} \tag{A.6}$$

Under the standard contingency, high rates are rewarded on the right and low rates on the left, so for a trained observer that has fully learnt the contingency,  $q(\text{High}, R) \rightarrow r_R$ ,  $q(\text{High}, L) \rightarrow 0$ ,  $q(\text{Low}, R) \rightarrow 0$ ,  $q(\text{Low}, L) \rightarrow r_L$ , with  $r_R$  and  $r_L$  being reward magnitudes for rightward and leftward actions. This simplifies the action values to:

$$\begin{aligned}
 Q(R) &= p(\text{High}|x_A, x_V)r_R \propto p_{\text{High}}(1 - \Phi(\mu_0; \mu_M, \sigma_M))r_R \\
 Q(L) &= p(\text{Low}|x_A, x_V)r_L \propto p_{\text{Low}}\Phi(\mu_0; \mu_M, \sigma_M)r_L
 \end{aligned} \tag{A.7}$$

The max-reward decision rule involves picking the action  $\hat{a}$  with the highest expected reward:

$$\begin{aligned}
\hat{a} &= \operatorname{argmax} Q(a) \\
\text{i.e. } \hat{a} = R &\iff Q(R) > Q(L) \\
\iff p_{High}(1 - \Phi(\mu_0; \mu_M, \sigma_M))r_R &> p_{Low}\Phi(\mu_0; \mu_M, \sigma_M)r_L & \text{(A.8)} \\
\iff \Phi(\mu_M; \mu_0, \sigma_M) &> \frac{1}{1 + \frac{p_{High}r_R}{p_{Low}r_L}} \\
\iff w_A x_A + w_V x_V &> \Phi^{-1}\left(\frac{1}{1 + \frac{p_{High}r_R}{p_{Low}r_L}}; \mu_0, (\sigma_A^{-2} + \sigma_V^{-2})^{-\frac{1}{2}}\right)
\end{aligned}$$

In the special case of equal rewards and uniform stimulus and category priors, this reduces to choosing right when the weighted sum of observations is to the right of the true category boundary, i.e.  $w_A x_A + w_V x_V > \mu_0$ . Note that this is a deterministic decision rule for any given observations  $x_A$  and  $x_V$ , however, since these are noisy and gaussian distributed around the true stimulus rate  $s_{true}$ , the likelihood of making a rightward decision is given by the cumulative gaussian function  $\Phi$ :

$$\begin{aligned}
\text{For } p_{High} = p_{Low}, r_R = r_L \\
p(\hat{a} = R|s) &= p(w_A x_A + w_V x_V > \mu_0|s) \\
&= \Phi(s_{true}; \mu_0, \sigma) & \text{(A.9)} \\
\sigma &= \begin{cases} \sigma_A \text{ on auditory trials} \\ \sigma_V \text{ on visual trials} \\ (\sigma_A^{-2} + \sigma_V^{-2})^{\frac{1}{2}} \text{ on multisensory trials} \end{cases}
\end{aligned}$$

We can measure this probability empirically through the psychometric curve. Fitting it with a two parameter cumulative gaussian function yields  $\mu$  and  $\sigma$  which can be compared to ideal observer predictions. The  $\sigma$  parameter is then taken to reflect sensory noise; and with

the assumption of uniform priors and equal rewards, the  $\mu$  parameter is taken to reflect the subjective category boundary. Although  $\mu$  should equal  $\mu_0$  for the ideal observer, in practice it is treated as a free parameter, and deviations of  $\mu$  from  $\mu_0$  could reflect any of three possible suboptimalities: 1) a subjective category boundary mismatched to the true one, 2) mismatched priors, or 3) unequal subjective rewards of the two actions.

### A.6.2 Inattention model

The traditional model for lapse rates assumes that on a fixed proportion of trials, the animal fails to pay attention to the stimulus, guessing randomly between the two actions. We can incorporate this suboptimality into the ideal observer above as follows: Let the probability of attending be  $p_{attend}$ . Then, on  $1 - p_{attend}$  fraction of trials, the animal does not attend to the stimulus (i.e. receives no evidence), effectively making  $\sigma_{sensory} \rightarrow \infty$  and giving rise to a posterior that is equal to the prior. On these trials, the animal may choose to maximize this prior (always picking the option that's more likely a-priori, guessing with 50-50 probability if both options are equally likely), or probability-match the prior (guessing in proportion to its prior). Let us call this guessing probability  $p_{bias}$ . Then, the probability of a rightward decision is given by marginalizing over the attentional state:

$$\begin{aligned}
 p(\hat{a} = R|s) &= p(\hat{a} = R|s, \text{attend})p(\text{attend}) + p(\hat{a} = R|s, \sim \text{attend})p(\sim \text{attend}) \\
 &= p(\hat{a} = R|s)p_{attend} + p_{bias}(1 - p_{attend})
 \end{aligned}
 \tag{A.10}$$

Comparing this with the standard 4-parameter sigmoid used in psychometric fitting, we obtain

$$\begin{aligned}
 p(\hat{a} = R|s_{true}) &= \gamma + (1 - \gamma - \lambda)\Phi(s_{true}; \mu_0, \sigma) \\
 \implies \gamma + \lambda &= p_{attend}, \quad \frac{\gamma}{\gamma + \lambda} = p_{bias}
 \end{aligned}
 \tag{A.11}$$

where  $\gamma$  and  $\lambda$  are the lower and upper asymptotes respectively, collectively known as ‘‘lapses’’. In this model, the sum of the two lapses depends on the probability of attending, which could be modulated in a bottom up fashion by the salience of the stimulus; their ratio depends on the guessing probability, which in turn depends on the observer’s priors and subjective rewards.

### A.6.3 Motor error model

Lapses can also occur if the observer doesn’t always pick the reward-maximizing or ‘‘exploit’’ decision. This might occur due to random errors in motor execution on a small fraction of trials given by  $\epsilon$ , or it might reflect a deliberate propensity to occasionally make random ‘‘exploratory’’ choices to gather information about rules and rewards. This is known as an  $\epsilon$ -greedy decision rule, where the observer chooses randomly (or according to  $p_{bias}$ ) on  $\epsilon$  fraction of trials. Both these models yield predictions similar to those of the inattention model:

$$\begin{aligned}
 p(\hat{a} = R|s) &= p(\hat{a} = R|s)(1 - \epsilon) + \epsilon p_{bias} \\
 \implies \gamma + \lambda &= \epsilon, \quad \frac{\gamma}{\gamma + \lambda} = p_{bias}
 \end{aligned}
 \tag{A.12}$$

### A.6.4 Exploration model

A more sophisticated form of exploration is the ‘‘softmax’’ decision rule, which explores options in proportion to their expected rewards, allowing for a balance between exploration and exploitation through the tuning of a parameter  $\beta$  known as inverse temperature. In particular, in conditions of greater uncertainty about rules or rewards, it is advantageous to be more exploratory and have a lower  $\beta$ . This strategy is known as Thompson sampling, and can be achieved by sampling from a belief distribution over expected rewards and maximizing with respect to the sample, reducing to a softmax rule whose  $\beta$  depends on the total uncertainty in expected reward (Gershman, 2018).

$$\begin{aligned} p(\hat{a} = R|Q(a)) &= \frac{\exp \beta Q(R)}{\exp \beta Q(L) + \exp \beta Q(R)} \\ &= \frac{1}{1 + \exp(-\beta(Q(R) - Q(L)))} \end{aligned} \quad (\text{A.13})$$

The proportion of rightward choices conditioned on the true stimulus rate is then obtained by marginalizing over the latent action values  $Q(a)$ , using the fact that the choice depends on  $s$  only through its effect on  $Q(a)$ , where  $\rho$  is the animal’s posterior belief in a high rate stimulus, i.e.  $\rho = p(c = High|x_A, x_V)$ .  $\rho$  is often referred to as the *belief state* in reinforcement learning problems involving partial observability such as our task.

$$\begin{aligned} p(\hat{a} = R|s) &= \int_{Q(a)} p(\hat{a} = R, Q(a)|s) dQ \\ &= \int_{Q(a)} p(\hat{a} = R|Q(a)) p(Q(a)|s) dQ \quad \because \hat{a} \perp s | Q(a) \\ &= \int_{\rho} \frac{1}{1 + \exp(-\beta(\rho(r_R + r_L) - r_L))} \frac{\mathcal{N}(\Phi^{-1}(1 - \rho, 0, \sigma_{post}), \mu_0 - s, \sigma_{post})}{\mathcal{N}(\Phi^{-1}(1 - \rho, 0, \sigma_{post}), 0, \sigma_{post})} d\rho \end{aligned} \quad (\text{A.14})$$

Since lapses are the asymptotic probabilities of the lesser rewarding action at extremely

easy stimulus rates, we can derive them from this expression by setting  $\rho \rightarrow 1$  or  $\rho \rightarrow 0$ .

This yields

$$\gamma = \frac{1}{1 + \exp(\beta r_L)}, \quad \lambda = \frac{1}{1 + \exp(\beta r_R)} \quad (\text{A.15})$$

Critically, in this model, the upper and lower lapses are dissociable, depending only on the rightward or leftward rewards, respectively. Such a softmax decision rule has been used to account for suboptimalities in value based decisions (Dayan and Daw, 2008), however it has not been used to account for lapses in perceptual decisions. Other suboptimal decision rules described in perceptual decisions, such as generalized probability matching or posterior sampling (Acerbi, Vijayakumar, and Wolpert, 2014; Drugowitsch et al., 2016; Ortega and Braun, 2013) amount to a softmax on log-posteriors or log-expected values, rather than on expected values, and do not produce lapses since in these decision rules, when the posterior probability goes to 1, so does the decision probability.

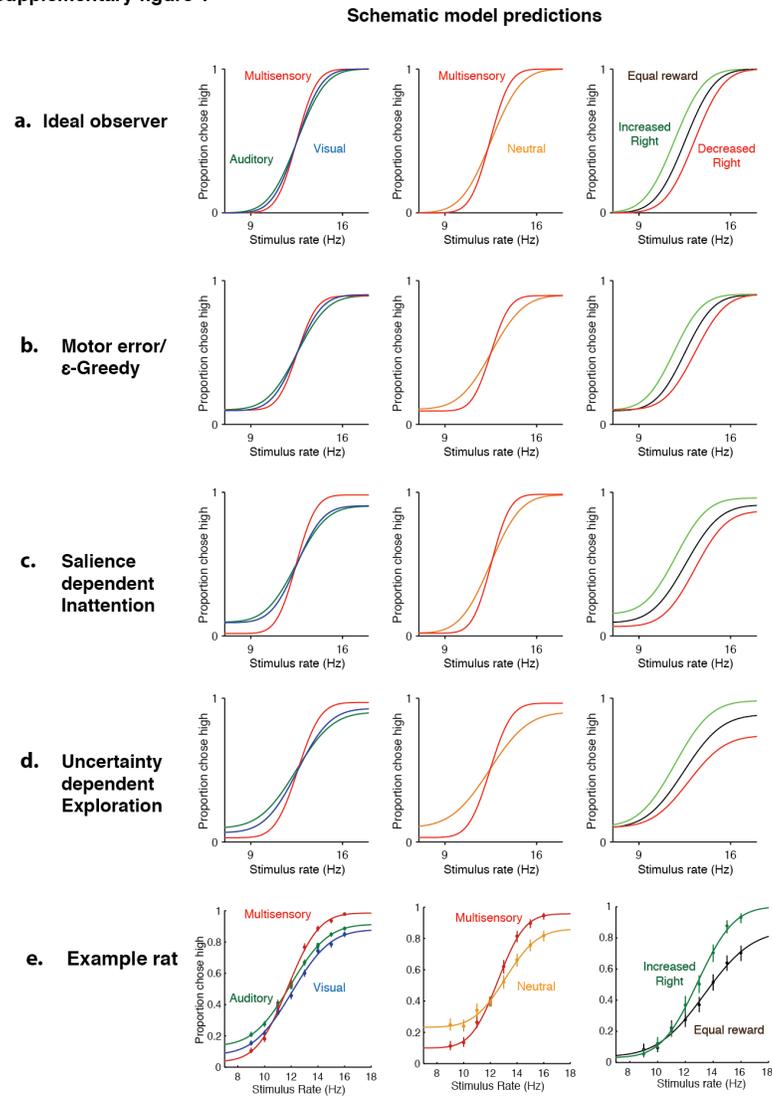
### A.6.5 Fitting and comparisons

Model fits were obtained from custom maximum likelihood fitting code using MATLAB's `fmincon`, by maximizing the marginal likelihood of rightward choices given the stimulus on each trial as computed from each model. Confidence intervals for fit parameters were generated using the hessian obtained from `fmincon`. Fits to multiple conditions were performed jointly, taking into account any linear or nonlinear (eg. optimality) constraints on parameters across conditions. Model comparisons were done using AIC and BIC.

## **Appendix B**

# **Supplementary figures**

Supplementary figure 1

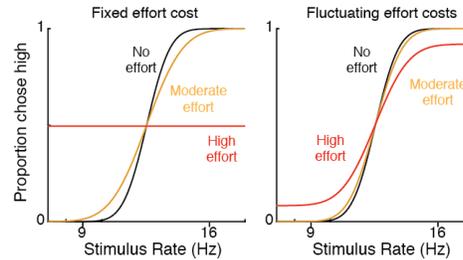


**FIGURE B.1: 1: Uncertainty-dependent exploration is the only model that accounts for behavioral data from all three manipulations**

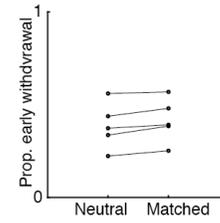
Columns: data/predictions for three experimental manipulations. Left: unisensory vs. multisensory. Middle: matched vs. neutral. Right: Asymmetric reward. a-d: Four candidate models. (a) Ideal observer model predicts no lapses and only changes in sensitivity/bias across conditions. (b) Fixed motor error model predicts a constant rate of lapses across conditions in addition to changes in sensitivity/bias predicted from the ideal observer. (c) Inattention model predicts that the overall lapse rate (sum of lapses on both sides) depends on the level of bottom-up attentional salience, allowing for different rates for unisensory and multisensory. It also predicts that the lapse rate on neutral trials should be equal to that on multisensory trials, and that manipulating rightward reward should affect both lapse rates. (d) Uncertainty-dependent exploration model predicts that overall lapse rate depends on the level of exploratoriness and hence uncertainty associated with that condition, allowing for different lapse rates on unisensory and multisensory trials. It also predicts that the lapse rate on neutral trials should be equal to that on auditory trials and manipulating rightward reward should only affect high rate lapses. (e) Data from an example rat

## Supplementary figure 2

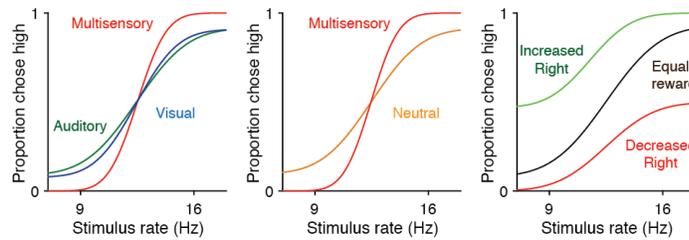
## a. Effort-dependent disengagement &amp; guessing



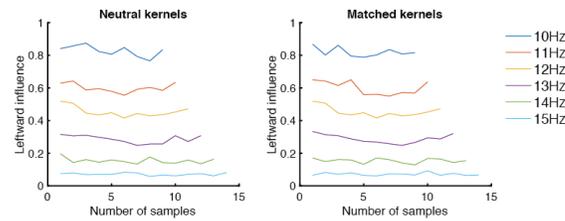
## b.



## c.



## d. Temporal inattention



## e. Variable precision

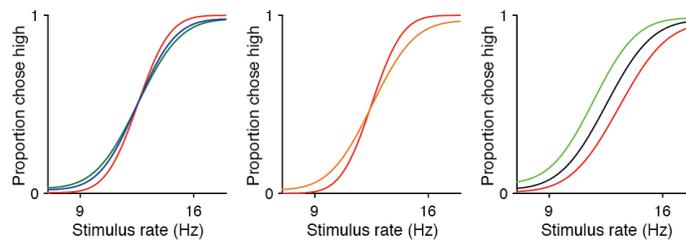
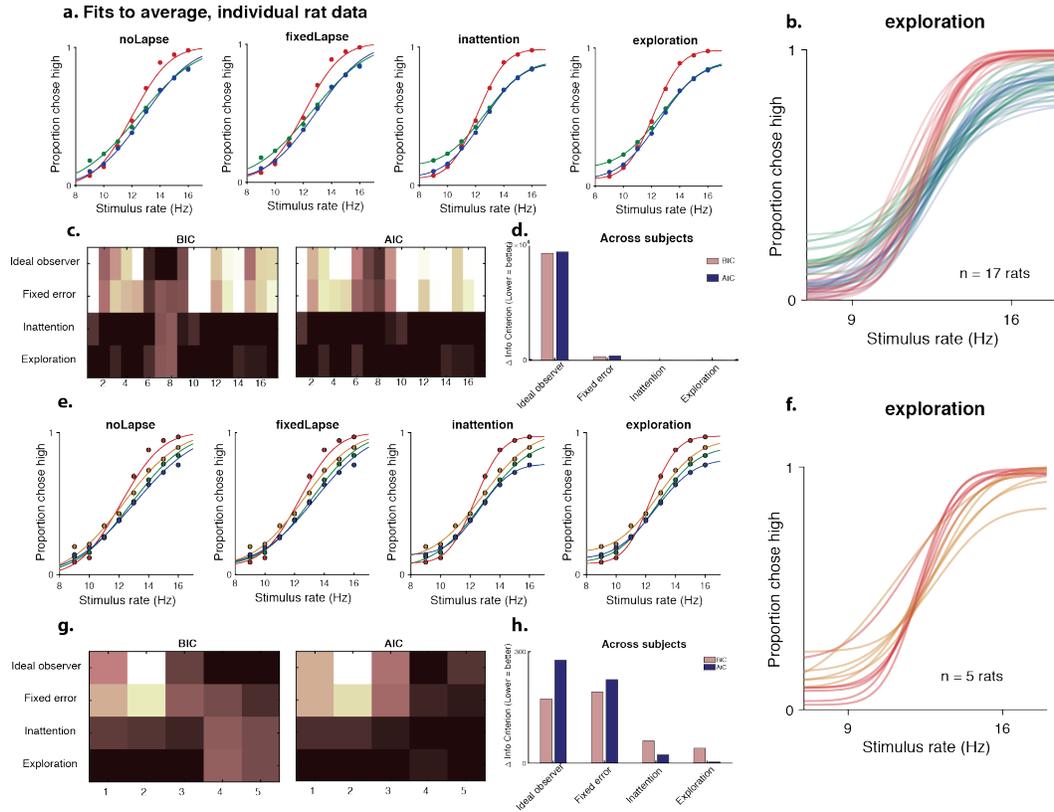


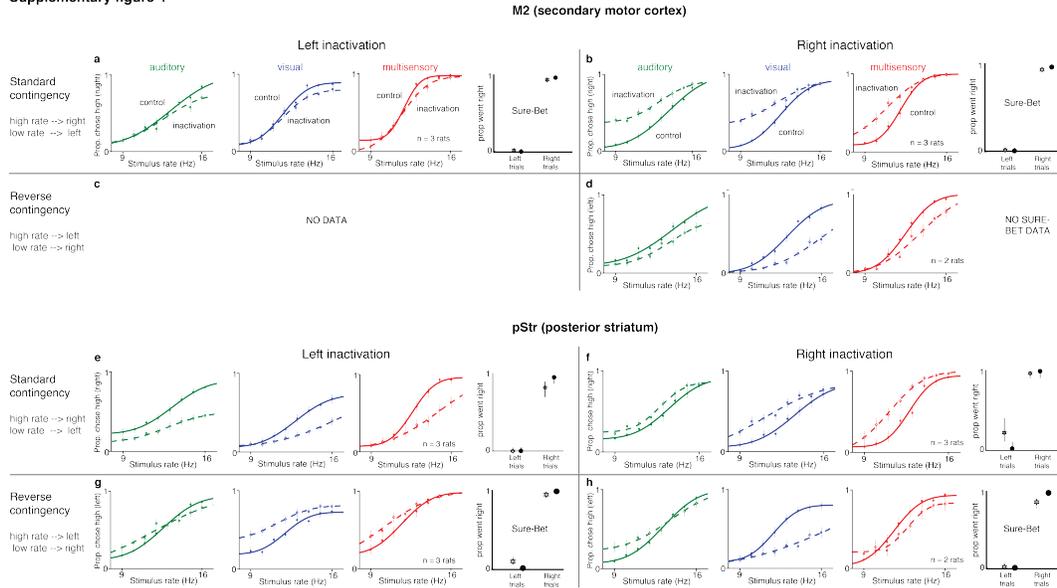
FIGURE B.2: **2: Alternative models of inattentional lapses.** Predictions of alternative models of lapses. (a) Effort-dependent disengagement model: In this model, there is an additional “effort” to being engaged in the task and an additional “random guessing” action. If the net payoff of engagement is not greater than the average value of a guess, then it guesses randomly. Such a model doesn’t produce lapses if the effort is fixed across trials (left), but could do so if the effort fluctuates from trial to trial (center). (b) Proportion of trials on which the animal withdrew prematurely doesn’t vary between matched and neutral trials, suggesting that rats are not disengaging preferentially on neutral trials. (c,e) Predictions of the effort-dependent disengagement model and Variable precision model respectively. These models accurately predicts increased lapses on unisensory (left panel, green/blue traces) and neutral trials (middle panel, orange trace) but fail to predict reward manipulation data. (d) Temporal inattention model: in this model, temporal weighting of evidence differs between matched and neutral trials. To test this, we compared psychophysical kernels on matched and neutral trials. The temporal dynamics of attention are unchanged between the two kinds of trials, arguing against temporal inattention.

Supplementary figure 3



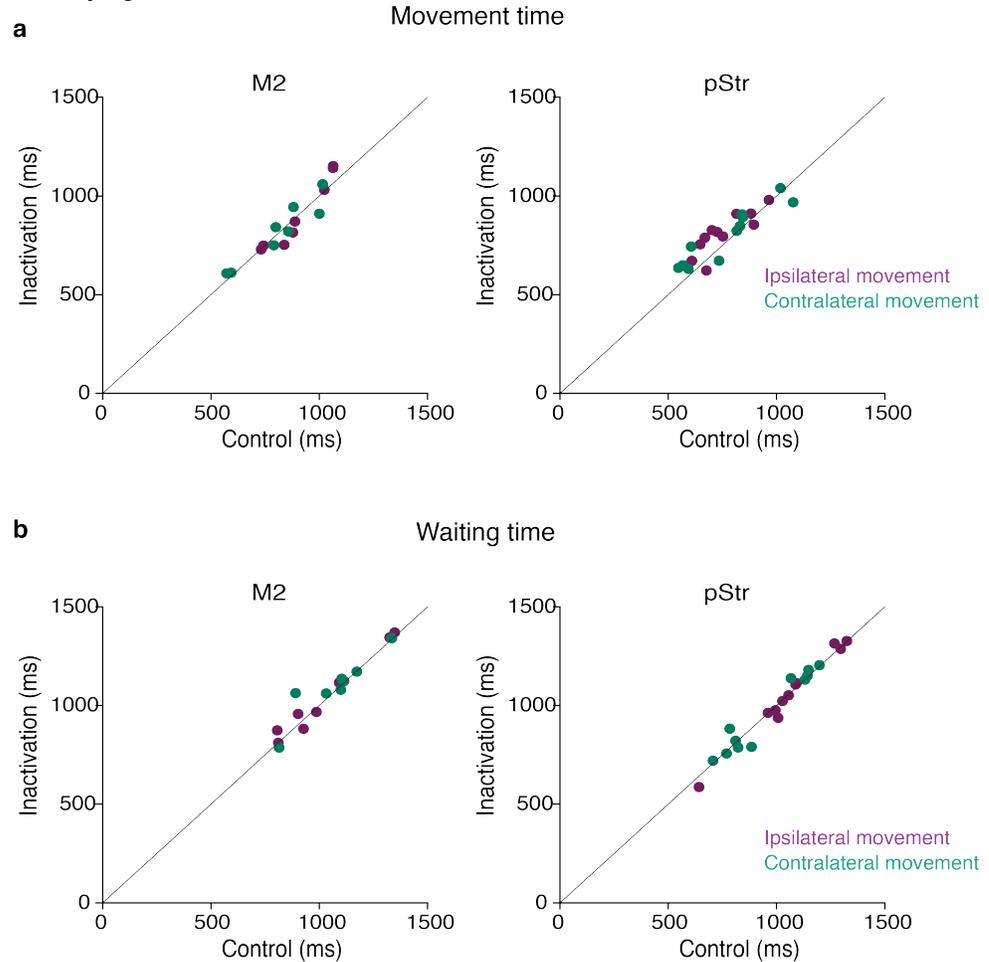
**FIGURE B.3: 3: Uncertainty guided exploration outperforms competing models for average and individual data** (a) Fits of the four models to average rat data on unisensory (blue-visual, green-auditory) and multisensory (red) trials. (b) Exploration model fits to unisensory and multisensory data for 17 individual animals (c) Model comparison for individual animals using BIC (left), AIC (right). Darker colors are lower BICs/AICs, denoting a better fit. (d) Summed model comparison metrics across animals, showing that inattention and exploration models fit the data equally well, and much better than the ideal observer or fixed error models. (e) Fits of the four models to average data including neutral trials (orange) provide a stronger test of the inattention model. (f) Exploration model fits to multisensory data including neutral trials for 5 individual animals (g) Model comparison for individual animals. (h) Summed model comparison metrics across animals shows that the uncertainty-guided exploration model performs better than other models.

Supplementary figure 4



**FIGURE B.4: 4: Unilateral inactivation of M2 or pStr biases performance ipsilaterally and increases contralateral lapses** Performance of the same rats shown in Figure 5b depicted as a function of the inactivated side (right or left) and the rate-contingency in which they were trained (standard or reverse). Standard contingency: high rate = go right, low rate = go left; reverse contingency: high rate = go left, low rate = go right. Each quadrant shows 4 plots: 3 psychometrics for rate discrimination trials and one for performance on sure-bet trials. auditory (green), visual (blue) and multisensory (red). (a)-(d) M2 inactivation. (e)-(h) pStr inactivation. (a), (d) Rats trained on the standard contingency and inactivated on the left hemisphere show increased lapses on the high rates (i.e., fewer rightward choices on high rates). No effect on sure-bet trials. (b), (f) Rats trained on the standard contingency and inactivated on the right hemisphere show increased lapses on the low rates (i.e., fewer leftward choices on low rates). No effect on sure-bet trials. (c), (g) Rats trained on the reverse contingency and inactivated on the left hemisphere show increased lapses on the low rates (i.e., fewer rightward choices on low rates). No effect on sure-bet trials. (d), (h) Rats trained on the reverse contingency and inactivated on the right hemisphere show increased lapses on the high rates (i.e., fewer leftward choices on high rates). No effect on sure-bet trials for pStr inactivated animals; no data for M2 inactivated animals.

## Supplementary figure 5



**FIGURE B.5: 10: No significant effect on movement parameters following muscimol inactivation** (a) Mean movement times from the center port to the side ports were not significantly different following muscimol inactivation of M2 (left;  $p = 0.9554$  for contralateral,  $0.9852$  for ipsilateral movements;  $n=5$  rats) or pStr (right;  $p = 0.6629$  for contra,  $p = 0.2615$  for ipsi,  $n=6$  rats). Control data on the abscissa is plotted against inactivation data on the ordinate. Purple, movement toward the side ipsilateral to the inactivation site; blue, movement toward the side contralateral to the inactivation site; Error bars (s.e.m.) are not visible because they were obscured by the markers in all cases. (b) Mean wait times in the center port were not significantly different following muscimol inactivation of M2 (left;  $p = 0.7612$  for contra,  $p = 0.8896$  for ipsi,  $n=5$  rats) or pStr (right;  $p = 0.9128$  for contra,  $p = 0.9412$  for ipsi,  $n=6$  rats). All  $p$ -values were computed from paired  $t$ -tests. Error bars (s.e.m.) are not visible because they were obscured by the markers in all cases.

# Bibliography

- Acerbi, Luigi, Sethu Vijayakumar, and Daniel M Wolpert (2014). “On the origins of sub-optimality in human probabilistic inference”. In: *PLoS computational biology* 10.6, e1003661.
- Barthas, Florent and Alex C Kwan (2017). “Secondary motor cortex- where sensory meets motor in the rodent frontal cortex”. In: *Trends in neurosciences* 40.3, pp. 181–193.
- Bays, Paul M, Raquel FG Catalao, and Masud Husain (2009). “The precision of visual working memory is set by allocation of a shared resource”. In: *Journal of vision* 9.10, pp. 7–7.
- Beck, Jeffrey M et al. (2012). “Not noisy, just wrong: the role of suboptimal inference in behavioral variability”. In: *Neuron* 74.1, pp. 30–39.
- Beeler, Jeff A et al. (2010). “Tonic dopamine modulates exploitation of reward learning”. In: *Frontiers in behavioral neuroscience* 4, p. 170.
- Bertolini, Giovanni et al. (2015). “Impaired tilt perception in Parkinsons disease- a central vestibular integration failure”. In: *PloS one* 10.4, e0124253.
- Bogacz, Rafal et al. (2006). “The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks.” In: *Psychological review* 113.4, p. 700.
- Brunton, Bingni W, Matthew M Botvinick, and Carlos D Brody (2013). “Rats and humans can optimally accumulate evidence for decision-making”. In: *Science* 340.6128, pp. 95–98.

- Burgess, Christopher P et al. (2017). “High-yield methods for accurate two-alternative visual psychophysics in head-fixed mice”. In: *Cell reports* 20.10, pp. 2513–2524.
- Busse, Laura et al. (2011). “The detection of visual contrast in the behaving mouse”. In: *Journal of Neuroscience* 31.31, pp. 11351–11361.
- Carandini, Matteo and Anne K Churchland (2013). “Probing perceptual decisions in rodents”. In: *Nature neuroscience* 16.7, p. 824.
- Cloherty, Shaun L et al. (2019). “Motion perception in the common marmoset”. In: *bioRxiv*, p. 522888.
- Cohen, Marlene R and John HR Maunsell (2009). “Attention improves performance primarily by reducing interneuronal correlations”. In: *Nature neuroscience* 12.12, p. 1594.
- Dayan, Peter and Nathaniel D Daw (2008). “Decision theory, reinforcement learning, and the brain”. In: *Cognitive, Affective, & Behavioral Neuroscience* 8.4, pp. 429–453.
- Dobs, Katharina, Wei Ji Ma, and Leila Reddy (2017). “Near-optimal integration of facial form and motion”. In: *Scientific reports* 7.1, p. 11002.
- Drugowitsch, Jan and Alexandre Pouget (2018). “Learning optimal decisions with confidence”. In: *bioRxiv*, p. 244269.
- Drugowitsch, Jan et al. (2016). “Computational precision of mental inference as critical source of human choice suboptimality”. In: *Neuron* 92.6, pp. 1398–1411.
- Erlich, Jeffrey C et al. (2015). “Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat”. In: *Elife* 4, e05457.
- Ernst, Marc O and Heinrich H Bulthoff (2004). “Merging the senses into a robust percept”. In: *Trends in cognitive sciences* 8.4, pp. 162–169.
- Findling, Charles et al. (2018). “Computational noise in reward-guided learning drives behavioral variability in volatile environments”. In: *bioRxiv*, p. 439885.
- Flesch, Timo et al. (2018). “Comparing continual task learning in minds and machines”. In: *Proceedings of the National Academy of Sciences* 115.44, E10313–E10322.

- Frank, Michael J et al. (2009). “Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation”. In: *Nature neuroscience* 12.8, p. 1062.
- Garrido, Marta I, Raymond J Dolan, and Maneesh Sahani (2011). “Surprise leads to noisier perceptual decisions”. In: *i-Perception* 2.2, pp. 112–120.
- Gershman, Samuel J (2015). “A unifying probabilistic view of associative learning”. In: *PLoS computational biology* 11.11, e1004567.
- (2018). “Deconstructing the human algorithms for exploration”. In: *Cognition* 173, pp. 34–42.
- Gershman, Samuel J, Eric J Horvitz, and Joshua B Tenenbaum (2015). “Computational rationality: A converging paradigm for intelligence in brains, minds, and machines”. In: *Science* 349.6245, pp. 273–278.
- Green, David M, John A Swets, et al. (1966). *Signal detection theory and psychophysics*. Vol. 1. Wiley New York.
- Gu, Yong, Dora E Angelaki, and Gregory C DeAngelis (2008). “Neural correlates of multi-sensory cue integration in macaque MSTd”. In: *Nature neuroscience* 11.10, p. 1201.
- Guo, Lan et al. (2018). “Stable representation of sounds in the posterior striatum during flexible auditory decisions”. In: *Nature communications* 9.1, p. 1534.
- Hanks, Timothy D and Christopher Summerfield (2017). “Perceptual decision making in rodents, monkeys, and humans”. In: *Neuron* 93.1, pp. 15–31.
- Hanks, Timothy D et al. (2015). “Distinct relationships of parietal and prefrontal cortices to evidence accumulation”. In: *Nature* 520.7546, p. 220.
- Hou, Han et al. (2018). “Neural correlates of optimal multisensory decision making”. In: *bioRxiv*, p. 480178.
- Jiang, Haiyan and Hyoung F Kim (2018). “Anatomical inputs from the sensory and value structures to the tail of the rat striatum”. In: *Frontiers in neuroanatomy* 12.
- Klein, Stanley A (2001). “Measuring, estimating, and understanding the psychometric function: A commentary”. In: *Perception & psychophysics* 63.8, pp. 1421–1455.

- Lak, Armin et al. (2018). “Dopaminergic and frontal signals for decisions guided by sensory evidence and reward value”. In: *bioRxiv*, p. 411413.
- Law, Chi-Tat and Joshua I Gold (2009). “Reinforcement learning can account for associative and perceptual learning on a visual-decision task”. In: *Nature neuroscience* 12.5, p. 655.
- Leblois, Arthur, Benjamin J Wendel, and David J Perkel (2010). “Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors”. In: *Journal of Neuroscience* 30.16, pp. 5730–5743.
- Lee, A Moses et al. (2015). “Between the primate and reptilian brain- rodent models demonstrate the role of corticostriatal circuits in decision making”. In: *Neuroscience* 296, pp. 66–74.
- Leike, Jan et al. (2016). “Thompson sampling is asymptotically optimal in general environments”. In: *arXiv preprint arXiv:1602.07905*.
- Licata, Angela M et al. (2017). “Posterior parietal cortex guides visual decisions in rats”. In: *Journal of Neuroscience* 37.19, pp. 4954–4966.
- Lieder, Falk and Thomas L Griffiths (2019). “Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources”. In: *Behavioral and Brain Sciences*, pp. 1–85.
- Lucas, Christopher G et al. (2014). “When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships”. In: *Cognition* 131.2, pp. 284–299.
- Mendonca, Andre G et al. (2018). “The impact of learning on perceptual decisions and its implication for speed-accuracy tradeoffs”. In: *bioRxiv*, p. 501858.
- Mihali, Andra et al. (2018). “A Low-Level Perceptual Correlate of Behavioral and Clinical Deficits in ADHD”. In: pp. 1–23.

- Nikbakht, Nader et al. (2018). “Supralinear and supramodal integration of visual and tactile signals in rats: psychophysics and neuronal mechanisms”. In: *Neuron* 97.3, pp. 626–639.
- Noel, Jean-Paul (2018). “Supra-optimality may emanate from suboptimality, and hence optimality is no benchmark in multisensory integration”. In: *Behavioral and Brain Sciences* 41.
- Nunez, Michael D et al. (2018). “The latency of a visual evoked potential tracks the onset of decision making”. In: *BioRxiv*, p. 275727.
- Odoemene, Onyekachi et al. (2018). “Visual evidence accumulation guides decision-making in unrestrained mice”. In: *Journal of Neuroscience* 38.47, pp. 10143–10155.
- Ortega, Pedro A and Daniel A Braun (2013). “Thermodynamics as a theory of decision-making with information-processing costs”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 469.2153, p. 20120683.
- Piet, Alex T et al. (2017). “Rat prefrontal cortex inactivations during decision making are explained by bistable attractor dynamics”. In: *Neural computation* 29.11, pp. 2861–2886.
- Pinto, Lucas et al. (2018). “An accumulation-of-evidence task using visual pulses for mice navigating in virtual reality”. In: *Frontiers in behavioral neuroscience* 12, p. 36.
- Prins, Nicolaas (2012). “The psychometric function: The lapse rate revisited”. In: *Journal of Vision* 12.6, pp. 25–25.
- Prins, Nicolaas and Frederick AA Kingdom (2018). “Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox”. In: *Frontiers in psychology* 9.
- Rahnev, Dobromir and Rachel N Denison (2018). “Suboptimality in perceptual decision making”. In: *Behavioral and Brain Sciences* 41.

- Raposo, David, Matthew T Kaufman, and Anne K Churchland (2014). “A category-free neural population supports evolving demands during decision-making”. In: *Nature neuroscience* 17.12, p. 1784.
- Raposo, David et al. (2012). “Multisensory decision-making in rats and humans”. In: *Journal of neuroscience* 32.11, pp. 3726–3735.
- Roy, Nicholas G et al. (2018). “Efficient inference for time-varying behavior during learning”. In: *Advances in Neural Information Processing Systems*, pp. 5700–5710.
- Scott, Benjamin B et al. (2015). “Sources of noise during accumulation of evidence in unrestrained and voluntarily head-restrained rats”. In: *Elife* 4, e11308.
- Shalom, Shir and Adam Zaidel (2018). “Better than optimal”. In: *Neuron* 97.3, pp. 484–487.
- Shen, Shan and Wei Ji Ma (2019). “Variable precision in visual perception.” In: *Psychological review* 126.1, p. 89.
- Sheppard, John P, David Raposo, and Anne K Churchland (2013). “Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans”. In: *Journal of vision* 13.6, pp. 4–4.
- Stocker, Alan A and Eero P Simoncelli (2006). “Noise characteristics and prior expectations in human visual speed perception”. In: *Nature neuroscience* 9.4, p. 578.
- Treutwein, Bernhard and Hans Strasburger (1999). “Fitting the psychometric function”. In: *Perception & psychophysics* 61.1, pp. 87–106.
- Wang, Lupeng et al. (2018). “Activation of striatal neurons causes a perceptual decision bias during visual change detection in mice”. In: *Neuron* 97.6, pp. 1369–1381.
- Wichmann, Felix A and N Jeremy Hill (2001). “The psychometric function: I. Fitting, sampling, and goodness of fit”. In: *Perception & psychophysics* 63.8, pp. 1293–1313.
- Yu, Angela J and Jonathan D Cohen (2009). “Sequential effects: superstition or rational behavior?” In: *Advances in neural information processing systems*, pp. 1873–1880.

- 
- Zatka-Haas, Peter et al. (2019). “Distinct contributions of mouse cortical areas to visual discrimination”. In: *bioRxiv*, p. 501627.
- Zhou, Baohua et al. (2018). “Chance, long tails, and inference in a non-Gaussian, Bayesian theory of vocal learning in songbirds”. In: *Proceedings of the National Academy of Sciences* 115.36, E8538–E8546.