The role of reinforcement learning in perceptual decision-making

André Gil Fernandes de Mendonça

Dissertation presented to obtain the
Ph.D. degree in Biology | Neuroscience

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
December, 2015

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

The role of reinforcement learning in perceptual decision-making

André Gil Fernandes de Mendonça

Dissertation presented to obtain the
Ph.D. degree in Biology | Neuroscience

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by:

Champalimaud Foundation

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

Oeiras,
December, 2015

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER /UNL

Knowledge Creation

The role of reinforcement learning in perceptual decision-making.

André Gil Fernandes de Mendonça

A Dissertation
Presented to the Faculty
of Universidade Nova de Lisboa
in Candidacy for the Degree
of Doctor of Philosophy

Zachary F. Mainen

December 2015

To Rita

Acknowledgments

{*I'm not good with acknowledgments. I always feel that I won't be able to find the right words for what I mean. Thus, I decided to take the words of others. I love music and dance. Consider this a mashed-up playlist, a medley of lyrics. Some songs remind me of the footnoted people, some are in there just because the lyrics for some reason match a story or a place, sometimes both, sometimes none. So, don't read too much in-between the lines as sometimes they're just meant to be silly. Also, I'm always scared that I might have forgotten someone. If I did, I'm sorry...*}

Everyday when I wake up, I thank the lord I'm Welsh.[1] He said, I locked you in this body, I meant it as a kind of trial. You can use it for a weapon, or to make some woman smile.[2] {*But, I asked,*} who'll be my role-model, now that my role-model is gone, gone?[3]

{*I was lucky enough to be part of a we*} and we'll never be royals, royals. It don't run in our blood, that kind of luxe just ain't for us. We crave a different kind of buzz.[4] And they said it changes when the sun goes down, yeah, they said it changes when the sun goes down.[5]

{*Much, much later*} she said it grieves me so to see you in such pain, I wish there was something I could do to make you smile again.[6] {*Oh! But you did!*}

{*Looking back, when my despair was too much, not all was lost as you gave me the space and*} you came to take us, all things go, all things go, to re-create us.[7] And if my parents are crying, then I'll dig a tunnel from my window to yours, yeah, a tunnel from my window to yours.[8] {*Truth*

[1] To my sister, Rute
[2] To my parents, Elisa and Vitoriano
[3] To my family, specially Álvaro, Gabi, Simão and Tómas
[4] To *elas*...
[5] ...and to *eles*, they know who *they* are
[6] To Ana
[7] To Zach
[8] To Maria

is} triangles are my favorite shape, three points where two lines meet, toe to toe, back to back.[9] {*But time is short and I came to realize:*} it's the final countdown, the final countdown, the final countdown.[10] {*"I have to finish this..."*} and never mind that noise you heard, it's just the beasts under your bed, in your closet, in your head.[11]

Sometimes I wonder if the world's so small, that we can never get away from the sprawl, living in the sprawl.[12] 'Cause they know, and so do I, the high road is hard to find, a detour to your new life.[13] {*I learned that you should*} blow steam in the face of the beast - "sky could fall down, wind could cry now. Look at me motherfucker, I smile!".[14] Hang on to the good days, I can lean on my friends, they help me going through hard times.[15] And if I made a fool, if I made a fool, if I made a fool, on the road, there's always this and if I'm sewn into submission I can still come home to this.[16]

{*Because*} I feel your whisper across the sea, I keep you with me in my heart, you make it easier when life gets hard.[17]

Bad mistakes - I've made a few. I've had my share of sand kicked in my face but I've come through.[18] But there's one sound, that no one knows: what does the fox say?[19]

We are the people that rule the world, a force running in every boy and girl, all rejoicing in the world.[20] {*And*} I try to laugh about it, cover it

[9] To Alex
[10] To Eric
[11] To Gil
[12] To Patrícia
[13] To Anna
[14] To Zaca
[15] To Rita Félix
[16] To Chico, Timon, Nico, Andreia, Sam, Rodrigo
[17] To Inês
[18] To Masa
[19] To the lab
[20] To Joe

all up with lies, I try and laugh about it.[21] Ooh, when all I was searching for was me. Keep your head up, keep your heart strong.[22] "Bamos lá cambada, todos à molhada que isto é futebol total."[23] Some of those that work forces are the same that burn crosses. Uggh! Killing in the name of![24]

Lose yourself to dance, lose yourself to dance, lose yourself to dance![25] 'Cause we are the champions, my friends, and we'll keep on fighting 'til the end. We, are the champions.[26]

And if I had the choice, yeah, I'd always wanna be there, those were the best days of my life.[27]

Decisions are made and not bought. But I thought this wouldn't hurt a lot. I guess not.[28]

[21] To Marta
[22] To Susana
[23] To Rui
[24] To Alfonso
[25] To Antonia
[26] To Pedro
[27] To INDP 2008
[28] To the whole CNP

Título

O papel da aprendizagem por reforço em sistemas de decisão perceptuais.

Resumo

A acumulação de evidências é uma componente importante nas tomadas de decisão perceptuais (*perceptual decision-making*, PDM) que permite aos organismos mitigar os efeitos de incerteza no ambiente através da combinação temporal de informação. Os modelos teóricos mais simples de acumulação de evidências têm sido bem sucedidos a descrever aspectos relacionados ao desempenho em tarefas psicofísicas, capturando a inter-dependência entre precisão e tempos de reacção (*reaction times*, RTs). No entanto, outros aspectos chave deste tipo de modelos permanecem substancialmente ambíguos.

Um fenómeno que ainda não é bem compreendido reside no facto de nem todas as decisões beneficiarem de acumulação de evidência na mesma escala de tempo. Por exemplo, ratos que executam uma tarefa de discriminação auditiva parecem integrar evidência durante mais de um segundo. Mas, perante uma tarefa de categorização de misturas de odores os benefícios de integrar por mais tempo desaparecem. Uma possível explicação é a de que os mecanismos de integração neurais são específicos à modalidade sensorial. Diferenças nas velocidades de precisão (*speed-accuracy tradeoff*, SAT) têm sido propostas como possíveis explicações para as diferenças observadas entre estudos extremamente semelhantes. Uma proposta alternativa é que as diferenças de tempo surgem a partir de diferenças nos requisitos computacionais. Dado que espécie, modalidade e a estrutura de tarefa têm variado entre os diversos estudos, distinguir entre todas as possibilidades, a partir dos dados existentes, é extremamente difícil.

Em modelos de acumulação de evidências a principal fonte de incerteza é estocacidade na evidência sensorial. Estas flutuações rápidas permitem explicar os benefícios de integração temporal. Outra fonte potencialmente importante de incerteza reside nas flutuações na taxa média de acumulação. Tais flutuações correspondem variabilidade no mapeamento de dados sensoriais numa determinada escolha (tentativa e erro). Esta variabilidade na categoria seria particularmente importante quando o mapeamento de estímulo a acção têm que ser aprendidas de novo, como é o caso numa tarefa de classificação de estímulos perceptuais.

Neste estudo, este problema foi abordado através da comparação da inter-dependência entre RT e exactidão em duas tarefas de decisão que eram idênticas excepto na natureza dos estímulos apresentados. A primeira tarefa consistiu na categorização de uma mistura de odores em que a dificuldade foi aumentada ao fazer com que os estímulos se situassem mais próximos de uma determinada categoria. A segunda tarefa consistiu na identificação de odores em que a dificuldade foi aumentada através da redução da concentração total dos odores. Verificou-se que a mudança de RT durante um determinado intervalo de precisão era diferente entre as duas tarefas. A nossa hipótese é que as flutuações na separação entre categorias reflectem uma forma de constante de aprendizagem por reforço (*reinforcement learning*, RL). De acordo com RL, as escolhas devem ser reguladas por um mecanismo baseado na expectativa de recompensas. RL é uma teoria normativa que postula que os agentes aprendem o valor de diferentes opções de escolha a partir da história dos resultados obtidos previamente. Se um mecanismo de RL é responsável pela diferença observada entre tarefas, então os padrões de escolha devem reflectir os erros na previsão de recompensa que são gerados por diferentes combinações de estímulo / escolha / resultado.

Descobrimos que os modelos de difusão padrão conseguem descrever eficientemente tanto o comportamento de categorização como identificação

mas apenas isoladamente, nunca em simultâneo. Foi necessário a introdução de um mecanismo de RL para poder reconciliar as duas tarefas. As flutuações na separação categórica entre os estímulos afectou principalmente o desempenho na tarefa de categorização, reduzindo assim um desempenho que originalmente era quase perfeito. Este resultado permite explicar a pequena diferença de RT observada nesta tarefa. Por último, RL previu uma dependência no histórico de escolha cujo padrão e magnitude foi posteriormente confirmando nos dados. Estes resultados suportam a ideia de que RL amplifica a variabilidade sensorial, produzindo uma fonte interna de variabilidade sem implicar estocacidade.

Abstract

Evidence accumulation is an important core component of perceptual decision-making that allows organisms to mitigate the effects of environmental uncertainty by combining information in time. Simple theoretical models of evidence accumulation have been successful in critical aspects of performance in psychophysical tasks, capturing the inter-dependence between accuracy and reaction time (RT). Yet substantial ambiguity remains concerning key features of this class of models.

One not well-understood phenomenon is that not all kinds of decisions appear to benefit from evidence accumulation over the same time scale. For example, rats performing an auditory discrimination task appear to integrate evidence over one second. But in an odor mixture categorization task rats fail to benefit from longer sampling. A possible explanation is that neural integration mechanisms are specific to a given sensory modality. Different speed-accuracy tradeoffs (SATs) have been proposed as another possible explanation for differences seen across similar studies. An alternative proposal is that differences in time arise from different computational requirements. Given that species, modality, task structure have all varied across past studies, distinguishing amongst these possibilities from existing data is difficult.

In models of evidence accumulation, the chief source of uncertainty is stochasticity in sensory evidence. These rapid fluctuations account for the benefits of temporal integration. A potentially important source of uncertainty is trial-by-trial fluctuations in the mean rate of accumulation. Such fluctuations would correspond to variability in the mapping of sensory data onto evidence for a particular choice. Such "boundary" variability would be particularly important when the mapping from stimulus to action must be learned de novo, such as in a categorization task.

We addressed this problem by comparing the dependence of RT and accuracy in two decision tasks that were identical except for the nature

of the presented stimuli. The first was an odor mixture categorization task in which difficulty was increased by making the stimuli closer to a category boundary. The second was an odor identification task in which the difficulty was increased by lowering concentration. We found that the RTs change over a given range of accuracy differed between the two tasks. We hypothesized that boundary fluctuations reflect a form of on-going reinforcement learning (RL). According to RL, choices should be driven by expected rewards. RL is a normative theory that posits that agents learn the expected values of different choice options from the history of their outcomes. If an RL mechanism is responsible, then these reward-dependent choice biases ought to exhibit specific patterns that would depend on the magnitude of the reward prediction errors generated by different stimulus/choice/outcome combinations.

We found that standard diffusion-to-bound models could fit well either categorization or identification task performance alone, but not simultaneously. Only when we included trial-by-trial updating of the category boundary due to ongoing RL could both data sets be fit with the same model. Fluctuations in category boundary primarily affected performance in the categorization task, reducing a nearly perfect performance. This explained the small RT change observed in this task. Critically, RL predicted a history-dependence of choice biases whose pattern and magnitude were closely matched to the data. These findings support the notion that RL amplifies sensory variability, producing an internal source of variability without implying stochasticity.

Author Contributions

André G. Mendonça (AGM), Maria I. Vicente (MIV) and Zachary F. Mainen (ZFM) designed the experiments in Chapter 2. AGM, ZFM and Alexandre Pouget designed the models presented from Chapter 3 to 5. MIV conducted the experiments shown in Chapter 2 with assistance from AGM. MIV and AGM analyzed the behavioral data in Chapters 2. AGM analyzed the data and implemented the models shown in Chapters 3 to 5. Gil Costa (GC) and ZFM designed the experiment in Chapter 6. GC conducted the experiment from Chapter 6. GC and AGM analyzed the data in Chapter 6. AGM implemented the GLM model presented in Chapter 6.

Financial Support

This work was developed in the context of the International Neuroscience Doctoral Programme (INDP) of the Champalimaud Research Programme, Champalimaud Center for the Unknown, Lisbon, Portugal. The project originally entitled "Attentional modulation of odor discrimination in rodents" and later named "The role of reinforcement learning in perceptual decision-making" was carried out at Instituto Gulbenkian de Ciência, Oeiras, Portugal and at the Champalimaud Research Programme, Champalimaud Center for the Unknown, Lisbon, Portugal, under the scientific supervision of Zachary Mainen, Ph.D, and under the guidance of the Thesis Committee composed by Marta Moita, Ph.D, and Joseph Paton, Ph.D. This work was supported by the fellowship SFRH / BD / 33938 / 2009 from Fundação para a Ciência e Tecnologia, Portugal.

Contents

List of Tables

List of Figures

Abbreviations

2RM ..... Two-Race accumulation Model

BIC ..... Bayesian Information Criterion

DDM ..... Drift-Diffusion Model

DDM+bias ... Drift-Diffusion Model with choice Bias

DM ..... Decision-Making

DV ..... Decision Variable

EDM ..... Economical Decision-Making

EOG ..... Electro-OlfactoGram

FOF ..... Frontal Orienting Fields

GLM ..... Generalized Linear Model

LIP ..... Lateral Intra-Parietal area

MC ..... Mitral Cell

ML ..... Matching Law

NS ..... Nervous System

OB ..... Olfactory Bulb

OFC ..... Orbito Frontal Cortex

OT ..... Olfactory Tubecule

PDM ..... Perceptual Decision-Making

PETH ..... Peri-Event Time Histogram

PIR ..... PIRiform cortex

PPC ..... Posterior Parietal Cortex

QM ..... Quantum Mechanics

RDK ..... Random-Dot Kinematogram

RL ..... Reinforcement Learning

RL-DDM ..... Reinforcement Learning Drift-Diffusion Model

RT ..... Reaction Time

SA ..... Sequential Analysis

SAT ..... Speed-Accuracy Tradeoff

SDT ..... Signal Detection Theory

VS ..... Ventral Striatum

VTA ..... Ventral Tegmental Area

Chapter 1

Introduction

"Every puzzle has an answer"
– Professor Layton, *Professor Layton and the Curious Village*

1.1 Chapter Summary

In this introduction we will address the topics of perceptual and economical decision-making. The goal of this introduction is to give the reader an overview and intuition of what are the challenges present while the brain tries to categorize perceptual stimuli.

To do so we divided this Chapter in 4 sections:

- **Randomness, noise and the brain** – in which we approach the subject of noise, uncertainty and random events and how the brain might cope with different sources of noise.
- **Sensory uncertainty** – a general review on the subject of sensory noise and its role in perceptual decision-making.
- **Actions, categories and the ever-changing environment** – an introduction on the subject of economical decision-making, with a particular focus on potentially extra sources of "noise".
- **Conceptual introduction and organization of the thesis** – where we focus our attention to the work presented in this thesis and introduce the conceptual idea behind it.

1.2 Introduction

The room is dark. You have no idea of what is this place, only darkness surrounds you. Slowly, an outline is drawn in the horizon and a beam of light propagates through the floor. You realize that there is a door, a door that is slightly ajar. You open it. You are almost blinded by the light.

As you cross the edge of the door you look around you. In front of you, a glade propagates into a forest full of strange colored and weird shaped trees, plants and animals. You have never experienced anything like this. Millions of photons bombard your retina; thousands of chemicals stimulate your olfactory receptors. As you walk through this forest you touch everything, anything that comes to reach. The textures of a leaf that grasped your attention feel funny, pointy while they look smooth. The patterns create a sensation that you're not familiar with. Your neurons are firing in ways that you have never felt before.

You see a strange fruit sitting in a brunch two feet away. You walk, but suddenly your legs are heavy, you take longer than you realize to reach this fruit. You feel massive, as if any small movement requires the strength of the world. You were so distracted as you left the room that you did not realize that even gravity has changed. Two feet, how hard can it be?

You finally reach this strange new fruit. It's shaped like a pyramid, something that you have never seen before. Its color sits within a spectrum that you don't easily define. It's a strange mixture of blue and green. You are tempted to say blue, but unsure. Let's say it's green-blue. "Green-blue fruit" has a texture that reminds you of an orange, so you peal it as you would peal an ordinary orange. The inside is sort of colorless; it looks like suspended water held back by small white filaments resembling a fishing net, shaping everything into a large pyramidal segment that fits

your hand perfectly. This is strange, you think to yourself. But you are so terribly thirsty, and so so hungry. You eat the strange fruit.

The effects are almost immediate, you feel massively rewarded; you have never felt anything like this. If words could describe your feelings, you would say that this fruit felt like fresh silk going down your throat. You want more and more of this fruit.

Or maybe the effects are just the opposite. You feel terribly sick after some minutes. You see everything move, rotations occur around you and you're the center axis. A massive bush that was sitting right next to you (*"was it always there!?"*) starts to morph into a face; a green and blue and black monster looks straight at you with his bloody eyes. You panic, what is this? What is happening to me? You try to run but you can't. The face dissolves and transforms itself into thousands of snakes that twist and twirl in your direction. As the first touches your skin you see that it was naught but a branch blown by wind.

You are now asked to remove your virtual-reality goggles. You now remember that this was all part of a simulation. But you will be asked to do it again.

Did you memorize the exact hue of that green-blue color? Will you pick it up again?

How did you create a story, a map from the sensory input that you were collecting? How did the photons hitting your retina create an impression in your brain that was categorizable? And more important, how did you categorize anything at all? How do you know what is green? What is blue? How do you separate between the two? What happened to you when you were hit by the effects of the fruit? Was reality really changing or was it you in your mind that misinterpreted that input? Does your brain generate the noise that made you perceive what you have perceived? Did you "decide" what you perceived? Or, more relevant perhaps, how does your brain "decide" for you?

This small piece of fiction is meant to illustrate the potential within our brains to extract meaning and categories from sensory input. But it also metaphorically depicts the fundamental question of perceptual decisions-making and that lives at the heart of this thesis: what modulates our percepts?

1.3 Randomness, noise and the brain

Randomness is defined as the lack of pattern or, more importantly for the case of neuroscience, lack of predictability in events. Events that are unintelligible in pattern (both spatially or temporally) are considered to be random. Randomness is considered to be a measurement of uncertainty in a particular outcome and has been applied to concepts of chance, probability and information theory (Bennett, 2009).

In the field of physics, the idea of random motion was fundamental for the development of statistical physics (Landau & Lifshitz, 2013). The incorporation of such concepts was paramount for the explanation of phenomena observed in thermodynamics and chemistry. Randomness is also key for the field of quantum mechanics (QM). Take the example of an unstable atom in a controlled environment. According to QM, its decay cannot be predicted efficiently, only the probability of it decaying in a given amount of time. QM works at the level of event probabilities and not at the resolution of individual outcomes (Feynman, Leighton, & Sands, 1963).

However, is randomness "truly" random? That is, are there events that are intrinsically "noisy" by nature and thus unpredictable even considering all variables? Albert Einstein, the father of Brownian motion (Einstein, 1905), did not believe that "God plays dice with the world" (Hermanns & Einstein, 1983). In fact, hidden variable theories reject the idea of nature containing truly random events (Einstein, Podolsky, &

Rosen, 1935). These theories posit that apparent random processes contain variables from statistical distributions that are working behind the scenes, and thus not immediately accessible or visible.

In biology, randomness has been described to exist in evolution. In particular genetic mutations have been thought to be random and an injector of variability upon which natural selection works on (Hastings, Lupski, Rosenberg, & Ira, 2009; Abby & Daubin, 2007). The environment would then work as the driving force behind deterministic characteristics arising (Klasson & Andersson, 2004). But is that always the case? Alternative mechanisms for non-random induced genetic variability in a population have been presented in the past (Wright, 2000; Martincorena, Seshasayee, & Luscombe, 2012).

In neuroscience, noise has been used to define unpredictable events that exist in the environment (Waiblinger, Brugger, Whitmire, Stanley, & Schwarz, 2015; Buzsáki, Peyrache, & Kubie, 2015), in human and animal behavior (Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979; de C. Hamilton, Jones, & Wolpert, 2004; Lum, Zhurov, Cropper, Weiss, & Brezina, 2005; Hooper, Guschlbauer, von Uckermann, & Buschges, 2006) and in neural activity (Deneve, Latham, & Pouget, 2001; Fitzpatrick, Batra, Stanford, & Kuwada, 1997; Kasamatsu, Polat, Pettet, & Norcia, 2001; Rolls & Deco, 2010; Shadlen, Britten, Newsome, & Movshon, 1996; Stocker & Simoncelli, 2006). Despite this, opposing ideas and theories have been brought forward that support (Faisal, Selen, & Wolpert, 2008) or challenge (Beck, Ma, Pitkow, Latham, & Pouget, 2012) the true nature of these sources of noise.

In any case, an open question of neuroscience is whether truly random sources of noise occur in the nervous system (Faisal et al., 2008; Rolls & Deco, 2010). To answer it is of unfathomable value as it fundamentally addresses hard questions regarding memory (Kandel, 2001), motor action generation (Orsborn & Carmena, 2013) or free-will (Libet, 1985;

Bengson, Kelley, Zhang, Wang, & Mangun, 2014). In this work, we aim to help shed light on some of these issues. We aim to understand the unpredictable/random/variable events in rodents performing a perceptual decision-making (PDM) task. In particular, we are interested in identifying what types of uncertainty does the nervous system (NS) have to cope with, and why or in which occasions they might be limiting behavior and respective performance.

1.3.1 Optimality

When studying the subject of noise and uncertainty one important concept that comes forward is optimality (Wald & Wolfowitz, 1949; Ernst & Banks, 2002). Behavior is considered optimal when it is limited merely by the level of noise/uncertainty in an environment or that was transmitted to the decision-maker (through noisy sensors for instance; Ash, 2012). It defines a good, if not the best, decision for a particular context (Wald & Wolfowitz, 1949; Yang & Shadlen, 2007; Ernst & Banks, 2002; Kording & Wolpert, 2004). However, what constitutes an optimal decision is in many ways a matter of disagreement between neuroscientists (see Summerfield & Tsetsos, 2015 for a review). In fact, different interpretations of the same behavioral result might originate strikingly different consequences (Summerfield & Tsetsos, 2012). Consider the example of Osborne, Lisberger, and Bialek (2005) in which primate subjects were asked to fixate and track visual targets. By analyzing the smooth-pursuit eye movements, Osborne et al. were able to find that the monkeys' behavior was optimal, conditional on the existence of constant small background noise during eye movements and fixations. The implication for this result is that the optimal solution is only existent if the brain is able to generate the extra noise needed to go from stimulus to behavior. Multiple hypothesis have been brought forward as a mechanism of how this injection of noise might occur (Faisal et al., 2008), in particular by considering noise build-up net-

works (Von Neumann, 1956; Laughlin & Sejnowski, 2003). However, a different interpretation for the same result is that monkeys are in fact inferring sub-optimally (Beck et al., 2012). In that particular study, Beck et al. suggest that all noise might be purely driven by sensors and amplified by choices that suffered deterministic (but suboptimal) approximations.

It is clear that individuals are not optimal in some particular cases (Yu & Cohen, 2008). The instructions given or the environmental set up might make the same subject oscillate from an observable optimal strategy to a non-optimal one (Summerfield & Tsetsos, 2012). In particular, a study in humans showed that subjects exposed to a rapidly changing perceptual environment were best explained by a non-optimal strategy (working memory rather than Bayesian inference; Summerfield, Behrens, & Koechlin, 2011). Considering that many results have showed that humans and monkeys can classify optimally visual information (Stocker & Simoncelli, 2006; Michel & Jacobs, 2008; Anderson, 1991; Ashby & Gott, 1988), one is led to conclude that the nature of the task at hand plays a significant role in understanding the strategies conducted by the NS. The gaps between optimal and non-optimal strategies might tell us more of how the brain copes with the external environment and help explain part of the observed variability of many tasks (Summerfield & Tsetsos, 2012).

Last but not least, if one considers the artificial setting of a controlled lab environment and its intrinsic difference with bona fide habitats, then one is led to the conclusion that these studies might help shed light on what are the significant evolutionary traits that emerged in the NS.

1.4 Sensory uncertainty

In PDM the main focus of research has been on the mechanisms by which observers categorize sensory signals. PDM tasks typically require subjects

to classify noisy sensory information. Here we explore how that sensory uncertainty has been defined and addressed in the field.

1.4.1 Perceptual decision-making

PDM is the process by which sensory information is used to guide behavior toward the external world. This involves gathering, evaluating and integrating information that was acquired through the senses. This information is then taken into account to produce judgments about the environment and conduct motor responses (Gold & Shadlen, 2007).

At any given point in time, the state of the world must be inferred based on noisy data provided by the sensory systems. This makes behavior critically dependent on the ability to quickly and accurately decide amongst the possible states. For instance, deciding whether or not a predator is present in a rich environment can dictate the survival chances of an animal. Various factors must be taken into account before committing to a decision and executing the appropriate behavioral response, including prior knowledge (discussed below). An important factor is the quality of the evidence derived from the sensory observations, then transformed into a decision variable (DV) that is interpreted by the decision layer/rule to produce a particular percept and choice (Gold & Shadlen, 2007). A conceptually simple and recurrently used approach is to consider a rule that places a decision criterion on the DV. The magnitude of the DV will then reflect the balance of support/opposition for a given choice, allowing the decision maker to achieve different goals; these might include accuracy, reward maximization or even targeting a particular decision time (Gold & Shadlen, 2007).

1.4.2 Signal detection theory

The study of perception and psychophysics has been a focus of decision theory since the XIX century (Fechner, 1948; Smith, 1994). Mathematical descriptions have been brought forward ever since, being signal detection theory (SDT) one of most historically relevant (Green & Swets, 1966).

SDT describes the process wherein inherent sensitivity of subjects to relevant stimuli is combined to generate choices, setting a framework to understand performance in perceptual tasks (Green & Swets, 1966). According to SDT, the decision-maker obtains an observation of noisy evidence from the stimulus, which gives rise to the DV that is then evaluated according to the decision rule. In simple binary decisions, the DV is typically related to the likelihood ratio of the different alternatives, and then compared to a given criterion. This criterion can also incorporate different priors (through Bayes rule, for instance) and value, allowing a flexible structure to hypothesize about PDM (Gold & Shadlen, 2007; Glimcher & Fehr, 2014).

1.4.3 Sequential analysis

SDT is focused on the nature of the decision before the DV has been completely evaluated. Sequential analysis (SA) is a natural extension to SDT that accommodates multiple pieces of evidence observed over time (Wald & Wolfowitz, 1949). The conceptual idea resides in that a decision-maker can benefit from multiple samples of a noisy distribution of variables that represent a stimulus. After each acquisition step, the DV is calculated from the evidence obtained up to that point and compared to the decision rule. This iterative process is then typically compared to a positive and negative criterion, each corresponding to a particular choice. Once the DV exceeds a criterion bound, a decision is made. The advantage of this

framework is that it allows predictions of how response times are generated (Edwards, 1965).

Several versions of sequential sampling models have been instantiated in the past (Gold & Shadlen, 2007; Luce, 1986; Townsend & Ashby, 1983; Vickers, 1970; Vickers, Carterette, & Friedman, 2014; Usher & McClelland, 2001; Ratcliff & Smith, 2004). A particular important instantiation are random walk models. In these models the DV is a cumulative sum of evidence over discrete time steps. If the evidence is the logarithm of the likelihood ratio, then this process corresponds to the statistically-optimal Sequential Probability Ratio Test (Wald & Wolfowitz, 1949). Instead, if the evidence is sampled from a Gaussian distribution in infinitesimal time steps, the process is termed diffusion with drift or bounded diffusion (Ratcliff, 1978).

While many models provide an account of either RT (Townsend & Ashby, 1983) or accuracy (Green & Swets, 1966), sequential sampling models relate shapes of RT distributions with probabilities of correct and incorrect responses, thereby explaining how RT and choice accuracy jointly vary as a function of the experimental conditions of interest. An important part of SA is that it allows quantification of the noise associated to psychophysical processes. Additionally, SA can be used to model and explain neurophysiological data (Gold & Shadlen, 2007; Sajda, Philiastides, Heekeren, & Ratcliff, 2011). For instance, recordings of neural activity in primates performing a random dot kinematogram (RDK) task have shown neural correlates resembling SA DVs (Roitman & Shadlen, 2002).

1.4.4 Accumulator models and Speed-accuracy tradeoffs

Evidence accumulation is an important core component of perceptual decision-making that allows organisms to mitigate the effects of environmental uncertainty by combining information in time (Roitman & Shadlen, 2002; Palmer, Huk, & Shadlen, 2005; Chittka, Dyer, Bock, &

Dornhaus, 2003; Histed, Carvalho, & Maunsell, 2012; Bowman, Kording, & Gottfried, 2012; Brunton, Botvinick, & Brody, 2013; Gold & Shadlen, 2007; Ratcliff & McKoon, 2008). Simple theoretical models of evidence accumulation based on a random walk-to-bound have been successful in critical aspects of the performance of psychophysical tasks, capturing the dependence of accuracy (psychometric) and reaction time (chronometric) functions. Key elements of these models have begun to be tested both by searching for neural activity corresponding to model variables (Kiani, Hanks, & Shadlen, 2008; Roitman & Shadlen, 2002; Hanks, Ditterich, & Shadlen, 2006; Erlich, Brunton, Duan, Hanks, & Brody, 2015; Hanks et al., 2015) and by the use of more sophisticated task design and modeling (Brunton et al., 2013; Zariwala, Kepecs, Uchida, Hirokawa, & Mainen, 2013). Yet substantial ambiguity remains concerning nearly all critical features of this class of models, including the basic mechanisms supporting integration, how a bound is determined and the origins of apparent randomness.

One widely observed but not well-understood phenomenon is that not all kinds of decisions appear to benefit from accumulation of evidence over the same time scale. For example, monkeys performing integration of random dot motion (Roitman & Shadlen, 2002) and rats performing a click train discrimination task (Brunton et al., 2013) appear to integrate evidence over one second. But rats performing an odor mixture categorization task fail to benefit from odor sampling beyond 200-300 ms (Uchida & Mainen, 2003; Zariwala et al., 2013). A possible explanation is that neural integration mechanisms are specific to a given species and sensory modality. However, even animals performing apparently similar odor-based show integration over very different time windows (Rinberg, Koulakov, & Gelperin, 2006b; Abraham et al., 2004). Motivation for speed vs. accuracy, or speed-accuracy tradeoff (SAT) (Palmer et al., 2005; Khan & Sobel, 2004; Hanks et al., 2006) has been proposed as a possible expla-

nation for differences seen across similar studies, although manipulation of motivation in one case failed to support this explanation (Zariwala et al., 2013). An alternative proposal is that differences in reaction time (RT) arise from different computational requirements of different tasks (Uchida, Kepecs, & Mainen, 2006; Zariwala et al., 2013; Summerfield & Tsetsos, 2012). Given that species, modality and task structure all vary across the different studies in question, distinguishing amongst these possibilities from existing data is difficult.

1.5 Actions, categories and the ever-changing environment

Although sensory information is ambiguous in PDM, a problem that arises from a simplistic approach is that a significant subset of these tasks present over-trained subjects that have developed a clear idea of how the sensory-to-action contingencies should work (Britten, Shadlen, Newsome, & Movshon, 1993). It is the identity of the stimulus, and thus, sensory uncertainty that is driving decisions in these type of tasks. However, what happens when these contingencies are not clear? How does a subject pick from a set of actions when perceptual categories and respective context might be changing on a trial-by-trial basis? Or when a subject *believes* the environment to be changing?

Here we explore how other sources of uncertainty (that also exist in PDM, see Busse et al., 2011 for an example) have been defined and addressed in the field of economic decision-making (EDM).

1.5.1 Matching law

Most behavioral decisions have the intrinsic goal of accomplishing a "reward". Food (Derusso et al., 2010), water (Uchida & Mainen, 2003), sex

(Nomoto & Lima, 2015), or social interactions (Marquez, Rennie, Costa, & Moita, 2015), have all been used as rewards for an animal. On the opposite scale, punishments have also been used as they generate negative action values that shun particular decisions (Paton, Belova, Morrison, & Salzman, 2006). A logical starting point is to consider that the net value of a particular decision is dictated by the needs that an agent wishes to satisfy. By this premise, an animal will choose to perform certain actions in detriment of others, so to maximize the rate of obtained reward - this is known as the matching law (ML).

The ML was first formulated by Herrnstein (1961) following an experiment with pigeons on concurrent variable interval schedules. Pigeons were presented with two buttons in a box, each with varying rates of food reward. The pigeons tended to peck the button that yielded the greater food reward more often than the low reward option. The ratio of their rates to the two buttons matched the ratio of reward rates on the two buttons. In operant conditioning, ML is then defined as the quantitative relationship between the relative rates of response and reinforcers in a concurrent reinforcement schedule. In the case of the pigeons, if the two response alternatives A and B are offered, the ratio of response rates to A and B equals the ratio of reinforcements yielded by each response:

$$\frac{R_A}{R_B} = \frac{r_A}{r_B} \tag{1.1}$$

being R the amount of responses for a particular side and r the reward frequency of that same option.

1.5.2 Reinforcement learning

Value is a concept that has been key in the study of behavioral decision making. Following rewards and avoiding punishment are extremely important not only for Herrnstein's pigeons but for the behavior of any animal.

However, ML does not address the issue of how these action values are accessed by the agent or even implemented in the brain.

When it comes to classical conditioning, reinforcement learning (RL) models have been very successful in describing EDM. RL emerged from the fields of experimental psychology (Rescorla & Wagner, 1972) and machine learning (Sutton & Barto, 1998), and it describes a mechanism for how the value of a particular stimulus or action is learned. Its basic premise is that values are updated by considering a prediction error (i.e., how surprising a given outcome is) and the weight that that particular error should have (learning rate). A simple version of these class of models is to consider the trial-by-trial learning rule known as the Rescorla-Wagner delta rule:

$$w \mapsto w + \alpha\delta\mu \tag{1.2}$$

where α is the learning rate, which can be interpreted as the associability of the stimulus, μ, with the reward; and w the weight that maps a stimulus to an action. The crucial term here is to consider the prediction error, $\delta = r - w\mu$, which dictates how far of predicting the reward, r, might the stimulus be (or not) and thus updated accordingly.

1.5.3 Value-based decision-making

In EDM tasks stimuli are usually unambiguous, and thus sensory uncertainty very much reduced (Sutton & Barto, 1998; Daw, Niv, & Dayan, 2005; Kable & Glimcher, 2007; Rushworth, Noonan, Boorman, Walton, & Behrens, 2011). Nevertheless, the nature of these tasks is still challenging as the value associated between the action options might drift or jump unpredictably across the experiment (Behrens, Woolrich, Walton, & Rushworth, 2007; Summerfield et al., 2011), within the course of a single session (Corrado, Sugrue, Seung, & Newsome, 2005) or even because the value of different choices are comparable and noisy (Kable & Glimcher,

2007). In these particular cases the identity of the stimulus was clear to the agent, but its relative value and prospective action were uncertain (Louie, Khaw, & Glimcher, 2013).

In most EDM tasks the choice process is usually modeled by assuming a greedy policy (the most valuable option available to the agent) or by a "softmax" function that permits some level of noise in the decision process. However, these models exist at a meta-process modeling level and lack process implementation and development (Lindland, Sindre, & Solvberg, 1994). Namely, they lack mechanisms that explain how a particular stimulus might be integrated, compared and classified by a neural network, and thus dictate the action to pursue. Fusing RL models with accumulator models might help shed light on some of these issues (Summerfield & Tsetsos, 2012).

1.5.4 Orbitofrontal cortex and ventral striatum

One important issue in both PDM and EDM is the neural implementation of these conceptual models, specially when considering the crosstalk between the two fields, as we propose to do here.

A potentially interesting candidate to look at is the orbitofrontal-cortex (OFC). The role of OFC has been investigated in olfactory-guided decision (Feierstein, Quirk, Uchida, Sosulski, & Mainen, 2006; Kepecs, Uchida, Zariwala, & Mainen, 2008). Neurons in the OFC were found to encode information that was dependent of trial timing (Feierstein et al., 2006). Before a decision was made, OFC neurons encoded information about stimuli, but not about choice. Later in a trial, when the animal was moving towards the choice port, OFC neurons encoded choice direction. After the animal reached the choice port, OFC neurons encoded information about goal properties, such as goal location and/or reward presence.

These findings are in concordance with the view of OFC as playing a central role in goal monitoring (Padoa-Schioppa & Assad, 2008; Schoenbaum, Takahashi, Liu, & McDannald, 2011; Takahashi et al., 2013). OFC activity was also found to be encoding decision confidence in rats performing an odor mixture categorization task (Kepecs, Uchida, Zariwala, & Mainen, 2008).

Evaluation, or performance monitoring, is necessary to analyze the efficacy or optimality of a decision with respect to its goals (Shadlen & Kiani, 2013). OFC is particularly important for reward-based behaviors when values are inferred, for instance using model-based RL algorithms (Daw & Doya, 2006). Additionally, OFC has been shown to keep track of absolute stimulus value (Kable & Glimcher, 2009) and to encode value in a "common currency" that allows comparison between variables that exist at different domains (Summerfield & Tsetsos, 2012).

OFC has been suggested as a direct player in voluntary choice via its interconnectivity with the ventral striatum (Kable & Glimcher, 2009). Ventral striatum (VS) has been shown to share a similarly important role in evaluation of performance (Botvinick, Niv, & Barto, 2009). This area was found to interact with OFC to guide optimal courses of action that ultimately lead to rewards (Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; McDannald, Lucantonio, Burke, Niv, & Schoenbaum, 2011; Simmons, Ravel, Shidara, & Richmond, 2007). Moreover, VS neurons also correlate with decision confidence in odor categorization (Costa, 2015). A cortico-striatal circuit involving OFC and VS could then be extremely relevant to evaluate decisions and optimize actions, while taking into account confidence estimates that arise from sensory inputs (Uchida et al., 2006).

1.6 Conceptual introduction and organization of the thesis

In theoretical models of evidence accumulation, the chief source of uncertainty is stochasticity in incoming sensory evidence, modeled as Gaussian white noise around the true mean evidence rate (Ratcliff, 1978; Ratcliff & Smith, 2004). It is this rapidly fluctuating noise that accounts for the benefits of temporal integration. The nature and implications of other sources of variability have also been considered in diffusion models (Ratcliff, 1978; Ratcliff & Smith, 2004; Ratcliff & McKoon, 2008; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Brunton et al., 2013; Fründ, Wichmann, & Macke, 2014), including variability in starting position (Mulder et al., 2012), variability in non-accumulation time (Ratcliff & Smith, 2004) and variability in threshold or "bound" (Ratcliff, 1978).

A potentially important source of uncertainty is trial-by-trial fluctuations in the mean rate of evidence accumulation. Such fluctuations would correspond to variability in the mapping of sensory data onto evidence for a particular choice direction (Gold, Law, Connolly, & Bennur, 2008; Beck et al., 2008). It has been hypothesized that such fluctuations would introduce errors that, unlike rapid fluctuations, could not be mitigated by temporal integration and would therefore curtail the benefits of evidence accumulation (Uchida et al., 2006; Zariwala et al., 2013). Such "boundary" (not to be confused with the stopping "bound" in accumulation models) variability might differentially affect different sort of decision tasks, being particularly important when the mapping from stimulus to action must be learned de novo, such as in a categorization task (Uchida et al., 2006; Zariwala et al., 2013).

Here, we addressed this problem by comparing the dependence of RT and accuracy on difficulty in two odor-guided decision tasks that were identical except for the nature of the stimuli. The first was an odor mix-

ture categorization task (Uchida & Mainen, 2003) in which the difficulty was increased by making the stimuli closer to a decision category boundary. The second was an odor identification task in which the difficulty was increased by lowering stimulus concentration. We found that the change in reaction times over a given range of accuracy differed markedly between the two tasks, despite being tested in the same subjects with all other task variables constant. We sought to obtain direct evidence that such boundary fluctuations limits perceptual decision performance and contributes to task differences in SAT. To do so, we considered the hypothesis that these boundary fluctuations reflect a form of on-going RL (Rescorla & Wagner, 1972; Sutton & Barto, 1998). Although perceptual choices are normally considered to be driven by sensory information, according to RL, choices should be driven by expected rewards. RL is a normative theory that posits that agents learn the expected values of different choice options from the history of their outcomes and that options are chosen so as to maximize those values (Rescorla & Wagner, 1972; Sutton & Barto, 1998). It is known that reward history can produce biases that effect perceptual decisions, contributing to a reduction in decision accuracy (Busse et al., 2011). If an RL mechanism is responsible, then these reward-dependent choice biases ought to exhibit specific patterns that would depend on the magnitude of the reward prediction errors generated by different stimulus/choice/outcome combinations.

We found that standard diffusion-to-bound models, without RL, could fit well either categorization or identification task performance alone, but not both simultaneously. Only when we included trial-by-trial updating of the category boundary due to ongoing reinforcement learning could both data sets be fit with the same model. Trial-to-trial fluctuations in category boundary primarily affected performance in the categorization task, reducing considerably what would have been nearly perfect performance. This explained the relatively small change in reaction time observed for this

task. Critically, the introduction of RL predicted a history-dependence of trial-by-trial choice biases whose specific pattern and magnitude were closely, qualitatively and quantitatively, matched to the data.

Lastly, by considering the involvement of both OFC (Kepecs, Uchida, Zariwala, & Mainen, 2008; Lak et al., 2014) and VS (Costa, 2015) in odor categorization, we devised a generalized liner model and applied it to previously acquired data (Costa, 2015) in search of correlates associated to changes predicted by our RL model. We found that both OFC and VS neuronal activity is modulated by the outcomes of previous trials. In particular, OFC showed modulated responses dependent on the categorical uncertainty of to the stimulus at hand.

These findings support the notion that RL amplifies sensory variability, producing an additional source of decision variability without implying stochastic internal processes (Beck et al., 2012).

To better present our findings we divided this thesis in 6 additional Chapters.

- **Chapter 2** – where the behavioral results that comprise this thesis are presented and the difference between odor identification and mixture categorization explained.

- **Chapter 3** – in which the issue of sensory uncertainty is addressed via the implementation of accumulator models, and why there is a need for an extra source of "noise".

- **Chapter 4** – in which we analyzed the effects of rewards and errors in on-going performance and local changes of choice-bias.

- **Chapter 5** – where we present the extended version of our accumulator model that incorporates an RL rule. We demonstrate that by taking into consideration assumptions of where sensory noise comes

from and how it might be amplified by an RL rule, we are able to not only fit the behavioral data but also predict an additional dataset.

- **Chapter 6** – in this chapter we tested for the presence of neural signatures of weight updating that were expected by our RL model. We saw both OFC and VS presented significant effects, although different in nature regarding stimulus uncertainty.

- **Chapter 7** – in this last section of the thesis we discuss the main findings, implications and future directions of this work.

Chapter 2

Odor identification and mixture categorization

"We are our choices"
– Helios, *Deus Ex*

2.1 Chapter Summary

In this Chapter we present the behavioral data that serve as the backbone of this thesis. The Chapter is divided in three sections:

- **Introduction** – a brief introduction on the subject of Speed-accuracy tradeoffs.
- **Animal subjects** – where behavioral methodology details such as differences in tasks, training and testing are explained.
- **Behavioral results** – in which the behavioral results and differences between odor identification and mixture categorization are shown.

2.2 Introduction

The quote "Fast is fine, but accuracy is everything" has been attributed to famous gambler and deputy sheriff Wyatt Earp in various contexts. Notoriously, it was replicated by American actor Kevin Costner in the movie of the same name (Kasdan, 1994). However, many historians believe that the sentence originated from a Greek student of Socrates, Xenophon. Regardless of the dispute, the fact that one lived in the XIX century while the other in 300 BC tells us that the interplay between accuracy and speed has been a philosophical point of discussion throughout the years.

Relationships between accuracy and speed of decision-making, or speed-accuracy tradeoffs (SATs), have been extensively studied in humans and other species including monkeys, rodents and insects (Gold & Shadlen, 2007; Luce, 1986; Palmer et al., 2005; Roitman & Shadlen, 2002; Uchida, Poo, & Haddad, 2014; Chittka et al., 2003; Chittka, Skorupski, & Raine, 2015; Uchida & Mainen, 2003; Abraham et al., 2004; Rinberg et al., 2006b; Histed et al., 2012; Zariwala et al., 2013; Bowman et al., 2012). In many ways Earp's quote seems to be true. Still, many other situations might favor fast decisions such as environments in which resources are scarce (competition) or contexts of predator presence (survival; Dawkins, 2004). It would only be natural to assume that many evolutionary pressures potentiated fine tuning of SATs (Gold & Shadlen, 2007; Dawkins, 2004).

However, the range of SATs observed varies widely across studies for reasons that are unclear. For example, reported increases in RT with increased difficulty of perceptual discrimination range from over 500 ms in humans (Palmer et al., 2005) and monkeys (Roitman & Shadlen, 2002) performing a RDK task, to 100 ms in mice performing a visual contrast detection task (Histed et al., 2012), to less than 30 ms in rats performing

an odor mixture discrimination task (Uchida & Mainen, 2003). It is not known what accounts for such different degrees of SAT observed across different studies.

Motivation for speed vs. accuracy is thought to be a key parameter affecting SAT (Khan & Sobel, 2004) and is a possible explanation for the differences observed across similar studies showing SAT of smaller (Uchida & Mainen, 2003) or larger (Abraham et al., 2004; Rinberg et al., 2006b) magnitudes. Two alternative possibilities are that longer SATs reflect neural mechanisms that are species-specific and/or sensory modality-specific. An additional explanation is that SAT differences arise from differences in the underlying computational requirements of different decision-making tasks (Zariwala et al., 2013). Given that species, modality, task structure all vary across the different studies in question, these possibilities are not distinguishable from existing data.

Our strategy was to compare SATs in two behavioral tasks that were identical except for the nature of the stimuli that gives rise to task difficulty. The first was a stimulus noise driven task in which the difficulty was increased by lowering stimulus concentration. We named this task "odor identification". The second was an odor mixture categorization task (Uchida & Mainen, 2003) in which the difficulty was increased by making the stimuli closer to a decision or category boundary. Thus, by having the same subjects performing two tasks that were different only for the set of stimuli, and by holding species, modality and motivation, we were in a condition that allowed us to test if SAT was dependent on the nature of the task. Our motivation was to create two extremely similar tasks that required different strategies and in that way explore and understand what dictates SAT.

Below, we will depict the odor-guided tasks that compose the core behavioral data analyzed in this thesis. We will also show the differences

that exist between the two tasks and that motivated the central topic of this study.

2.3 Animal subjects

Four Long Evans rats (200-250 g at the start of training) were trained and tested in accordance with European Union Directive 86/609/EEC and approved by Direcção-Geral de Veterinária (DGV) of Portugal. Rats were trained and tested on three different tasks: (1) a two-alternative choice odor identification task; (2) a two-alternative choice odor mixture categorization task (Uchida & Mainen, 2003); and (3) a two-alternative choice "odor mixture identification" task. The same rats performed all three tasks that differ on the nature of the presented stimulus while all other task variables were held constant (Figure 2.1). Rats were pair-housed and maintained on a normal 12 hr light/dark cycle and tested during the daylight period. Rats were allowed free access to food but were water-restricted. Water was available during the behavioral session and for 20 minutes after the session at a random time as well as on non-training days. Water availability was adjusted to ensure animals maintained no less than 85% of ad libitum weight at any time.

2.3.1 Training and testing apparatus

The behavioral apparatus for the task was designed by Z.F.M. in collaboration with M. Recchia (Island Motion Corporation, Tappan, NY). The behavioral control system (BControl) was developed by Z.F.M, C. Brody (Princeton University) in collaboration with A. Zador (Cold Spring Harbor Laboratory). The behavioral setup consisted of a box (27×36 cm) with a panel containing three conical ports (2.5 cm diameter, 1 cm depth; Uchida & Mainen, 2003). Each port was equipped with an infrared photodiode/phototransistor pair that registered a digital signal when the rats

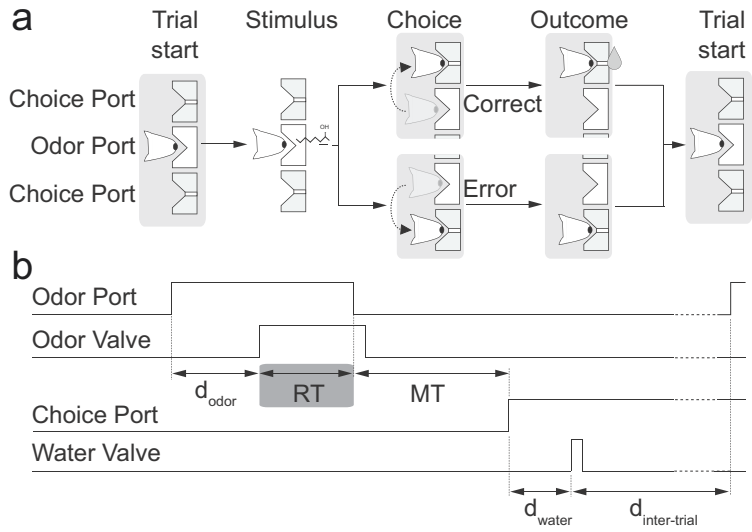

Figure 2.1. Two-alternative odor choice task. **(a)** Rats were trained in a behavioral box to signal a choice between left and right port after sampling a central odor port. The sequence of events is illustrated using a schematic of the ports and the position of the snout of the rats. **(b)** Illustration of the timing of events in a typical trial. Nose port photodiode and valve command signals are shown (thick lines). A trial is initialized after a rat pokes into a central odor port. After a randomized delay d_{odor} a pure odor or a mixture of odors is presented, dependent of the task at hand. The rat can sample freely and respond by moving into a choice port in order to get a water reward. Each of these ports is associated to one of two odors – odor A ((R)-()-2-Octanol) and odor B ((S)-(+)-2-Octanol). Highlighted by the grey box, reaction time (RT) is the amount of time the rats spend in the central odor port.

snout was introduced into the port ("nose poke"), allowing us to determine the position of the animal during the task with high temporal precision. Odors were delivered from the center port and water from the left and right ports. Odor delivery was controlled by a custom made olfactometer (Uchida & Mainen, 2003) designed by Z.F.M in collaboration with

M. Recchia. During training and testing the rats alternated between two different boxes.

The test odors were S-(+) and R-(-) stereoisomers of 2-octanol, chosen for their identical vapor pressures and similar intensities (Uchida & Mainen, 2003; Taniguchi, Kashiwayanagi, & Kurihara, 1992; Pierce, Zeng, Aronov, Preti, & Wysocki, 1995; Laska, Psychologie, München, & München, 2004). In the odor identification task, difficulty was manipulated by using different concentrations of pure odors, ranging from 10^{-4} to 10^{-1} (v/v) (Figure 2.2a). The different concentrations were produced by serial liquid dilution using an odorless carrier, propylene glycol (1,2-propanediol). In the odor mixture categorization task, we used binary mixtures of these two odorants at different ratios, with the sum held constant: 0/100, 20/80, 32/68, 44/56 and their complements (100/0, etc.; Figure 2.2b). Difficulty was determined by the distance of the mixtures to the category boundary (50/50), denoted as "mixture contrast" (e.g. 80/20 and 20/80 stimuli correspond to 60% mixture contrast). Choices were rewarded at the left choice port for odorant A (identification task; Figure 2.2c,d) or for mixtures $A/B > 50/50$ (categorization task; Figure 2.2e,f) and at the right choice port for odorant B (identification task) or for mixtures $A/B < 50/50$ (categorization task). In both tasks, the set of eight stimuli were randomly interleaved within the session. During testing, the probability of each stimulus being selected was the same.

We define as "mixture categorization task" sessions in which mixtures with a total odor concentration of 10^{-1} (v/v) were used. For the "odor mixture identification", we used the same mixture contrasts with total concentrations ranging from 10^{-1} to 10^{-4} prepared using the diluted odorants used for the identification task (Figure 2.5a). In each session, four different mixture pairs were pseudo-randomly selected from the total set of 32 stimuli (8 contrasts at 4 different total concentrations).

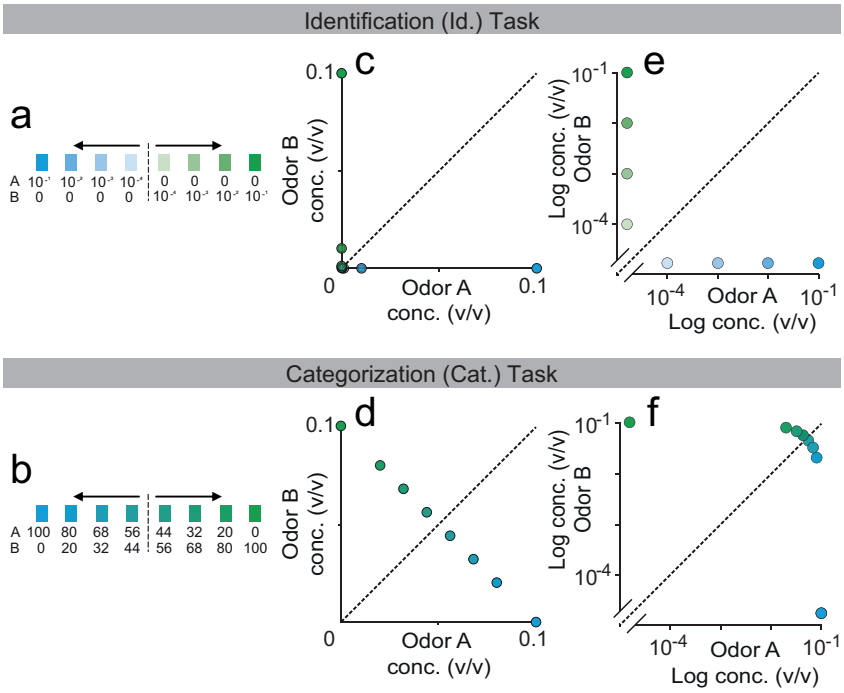

Figure 2.2. Stimulus design and task differences. **(a,b)** In the odor detection task, the odorants were presented independently at concentrations ranging 10^{-1} to 10^{-4} (v/v) and sides rewarded accordingly (a). For the mixture categorization task, the two odorants were mixed in different ratios presented at a fixed total concentration of 10^{-1}, and rats were rewarded according to the majority component (b). Each dot represents one of the 8 stimuli presented for each task. **(c,d)** Linear space representation of all stimuli presented in identification (c) and categorization task (d). All stimuli above the dashed identity line indicate a right side choice, and all stimuli below a left choice. **(e,f)** Same as (c,d) in a logarithmic scale.

For all the different experiments, four of the eight stimuli presented in each session were rewarded on the left ($A > B$) and the other four were rewarded on the right ($A < B$). Each stimulus was presented with equal probability and corresponded to a different filter in the manifold.

The training sequence consisted of: (I) handling (2 sessions); (II) water port training (1 session); (III) odor port training, in which a nose poke at the odor sampling port was required before water was available at the choice port. The required center poke duration was increased from 0 to 300 ms (4 - 8 sessions); (IV) introduction of test odors at a concentration of 10^{-1}, rewarded at left and right choice ports according to the identity of the odor presented (1 - 5 sessions); (V) introduction of increasingly lower concentrations (5 - 10 sessions); (VI) training on odor identification task (10 - 20 sessions); (VII) testing on odor identification task (14 - 16 sessions); (VIII) training on mixture categorization task (10 - 20 sessions); (IX) testing on mixture categorization task (14 - 15 sessions); (X) testing on mixture identification task (12 - 27 sessions) (Figure 2.3).

During training, in phases VI and VIII, we used adaptive algorithms to adjust the difficulty and to minimize bias of the animals. We computed an online estimate of bias:

$$b_t = (1 - \tau)C_t + \tau b_{t-1} \tag{2.1}$$

where b_t is the estimated bias in the current trial, b_{t-1} is the estimated bias in the previous trial, C_t is the choice of the current trial (0 if right, 1 if left) and τ is the decay rate ($\tau = 0.95$ in our experiments). The probability, p, of being presented with a right-side rewarded odor was adjusted to counteract the measured bias using:

$$p_R = 1 - \frac{1}{1 + e^{(b_t - b_0)/\gamma}} \tag{2.2}$$

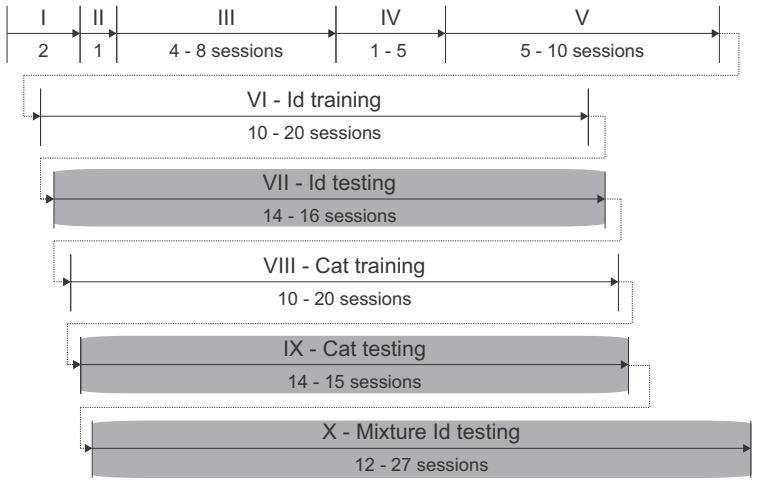

Figure 2.3. Tasks training and testing time-line. (I) Handling (2 sessions). (II) Water port training (1 session). (III) Odor port training (4 - 8 sessions). (IV) Pure odor training (1 - 5 sessions). (V) Introduction of lower concentrations (5 - 10 sessions). (VI) Odor identification training (10 - 20 sessions). (VII) Odor identification testing (14 - 16 sessions). (VIII) Mixture categorization training (10 - 20 sessions); (IX) Mixture categorization testing (14 - 15 sessions). (X) Mixture identification testing (12 - 27 sessions). Each session represents a different day. Grey boxes highlight the behaviorial data presented in Chapter 2 to 5.

where b_0 is the target bias (set to 0.5), and γ (set to 0.25) describes the degree of non-linearity.

Analogously, the probability of a given stimulus difficulty was dependent on the performance of the animal, i.e., the relative probability of difficult stimuli was set to increase with performance. Performance was calculated in an analogous way as Equation 2.1 at the current trial but c_t became r_t – the outcome of the current trial (0 if error, 1 if correct). A difficulty parameter, δ, was adjusted as a function of the performance,

$$\delta_{t+1} = -1 + \frac{2}{e^{(p_t - p_0)/\gamma}} \tag{2.3}$$

where p_0 is the target performance (set to 0.95). The probability of each stimulus difficulty, φ, was drawn from a geometric cumulative distribution function (GEOCDF, Matlab):

$$\varphi_{t+1} = \frac{1 - GEOCDF\big(i, |\delta_{t+1}|\big)}{\sum_{j=1}^{N} 1 - GEOCDF\big(j, |\delta_{t+1}|\big)} \tag{2.4}$$

where N is the number of stimulus difficulties in the session, and takes a value from 2 to 4 (when $N = 1$, i.e. only one stimulus difficulty, this algorithm is not needed); i corresponds to the stimulus difficulty and is an integer from 1 to 4 (when $\delta > 0$, the value 1 corresponds to the easiest stimuli and 4 to the most difficult one, and vice-versa when $\delta < 0$). In this way, when $|\delta|$ is close to 0, corresponding to an average performance close to 0.95, the distribution of stimuli was close to uniform (i.e. all difficulties are equally likely to be presented). When performance is greater, then the relative probability of difficult trials increased; conversely, when the performance is lower, the relative probability of difficult trials decreased. Training phases VI and VIII were interrupted for both tasks when number of stimulus difficulties $N = 4$ and difficulty parameter δ stabilized on a session-by-session basis.

Each rat performed one session of 90 – 120 minutes per day (250 - 400 trials), 5 days per week for a period of ∼120 weeks. During testing, the adapting algorithms were turned off and each task was tested independently.

2.3.2 Reaction time paradigm

The timing of task events are illustrated in Figure 2.1. Rats initiated a trial by entering the central odor-sampling port, which triggered the delivery of an odor with delay (d_{odor}) drawn from a uniform distribution with a range of [0.3, 0.6] s. The odor was available for up to 1 s after odor

onset. Rats could exit from the odor port at any time after odor valve opening, and make a movement to either of the two reward ports. Trials in which the rat left the odor sampling port before odor valve opening ($\sim$4% of trials) or before a minimum odor sampling time of 100 ms had elapsed ($\sim$1% of trials) were considered invalid. Odor delivery was terminated as soon as the rat exited the odor port. Reaction time (the odor sampling duration) was calculated as the difference between odor valve actuation until odor port exit (Figure 2.1b) minus the delay from valve opening to odor reaching the nose. This delay was measured with a photo ionization detector (mini-PID, Aurora Scientific, Inc) and had a value of 53 ms.

Reward was available for correct choices for up to 4 s after the rat left the odor sampling port. For correct trials, water was delivered from gravity-fed reservoirs regulated by solenoid valves after the rat entered the choice port, with a delay (d_{water}) drawn from a uniform distribution with a range of [0.1, 0.3] s. Trials in which the rat failed to respond to one of the two choice ports within the reward availability period (0.5% of trials) were also considered invalid. Reward amount (w_{rew}), determined by valve opening duration, was set to 0.024 ml and calibrated regularly. A new trial was initiated when the rat entered odor port, as long as a minimum interval ($d_{inter-trial}$), of 4 s from water delivery, had elapsed. Error choices resulted in water omission and a "time-out" penalty of 4 s added to $d_{inter-trial}$. Behavioral accuracy was defined as the number of correct choices over the total number of correct and incorrect choices. Invalid trials (in total 5.8 $\pm$ 0.8% of trials, mean $\pm$ SEM, n = 4 rats) were not included in the calculation of performance accuracy or reaction times.

2.3.3 Statistical and behavioral analysis

All the behavioral, statistical analysis, fittings and computational simulations were performed in Matlab®. For more details please refer to Vicente (2015).

2.4 Behavioral results

2.4.1 Odor identification versus odor categorization

We trained Long Evans rats on two different two-alternative choice olfactory reaction time tasks that were similar except for the stimulus concentrations (Figure 2.1).

In the first task, "odor identification", a single odor was presented at any given trial and we manipulated difficulty by diluting odors over a range of 3 log steps (1000-fold in liquid; Figure 2.2a). The absolute concentration of the odor determines the difficulty. In the second task, "odor categorization", mixtures of two odors were presented at a fixed total concentration but with varying ratios (Uchida & Mainen, 2003) (Figure 2.2b). The distance of the stimulus to the category boundary (50/50 iso-concentration line), termed "mixture contrast" (e.g., 56/44 and 44/56 stimuli correspond to 12% mixture contrast), determined the difficulty of a given trial, with lower contrasts corresponding to more difficult trials. Note the easiest odor pairs (10^{-1} dilution and 100% contrast) were identical between the two tasks. In a given session, eight randomly interleaved stimuli from one of the two tasks were presented. Critically, to ensure that any differences in performance were due to the manipulated stimulus parameters only, all comparisons were done using the same rats performing the two tasks on different days with all other task variables being held constant.

We quantified performance using accuracy (fraction of correct trials) and odor sampling duration, a measure for RT (Uchida & Mainen, 2003; Zariwala et al., 2013; Figure 2.1b). We observed that rats performing the two tasks showed marked differences in RT changes across stimuli for which there was a similar change in accuracy (Figure 2.4). For the identification task, RTs increased substantially (108 ± 28 ms; mean $\pm$ SEM, $n = 4$ rats; $F(3, 12) = 8.63$, $P < 0.01$; Figure 2.4d). In contrast, for the same animals performing the categorization task, RT showed a small

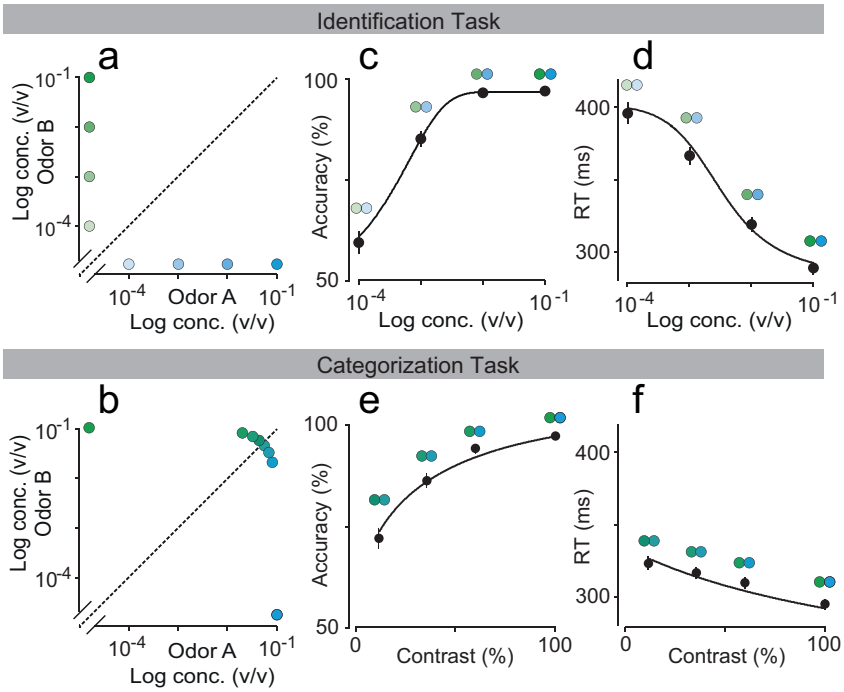

Figure 2.4. Comparison between odor identification and mixture categorization tasks. **(a,b)** Stimulus space in logarithmic scale for identification (a) and categorization (b) tasks. **(c,d)** Mean accuracy (c) and mean reaction time (d) for the identification task plotted as a function of odor concentration. **(e,f)** Mean accuracy (e) and mean reaction time (f) for the categorization task plotted as a function of mixture contrast (i.e. the absolute percent difference between the two odors). Error bars are mean ± SEM over trials and rats. Dots are presented as to help parse between stimulus space and psych- and chrono-metric curves. Solid lines depict the obtained fits for RL-DDM model (Chapter 5).

and non-significant increase from the easiest to the most difficult stimuli (34 ± 17 ms; $F(3, 12) = 1.62$, $P > 0.2$, ANOVA; Figure 2.4f).

2.4.2 Odor mixture identification task

Motivational variables can modulate performance and reaction time in perceptual tasks. For example, variables like reward rate (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012) or emphasis for accuracy vs. speed (Palmer et al., 2005; Hanks et al., 2006) can have an effect on observed SATs, by modulating decision criteria. Because identification and categorization tasks were run in separate sessions, we also considered the possibility that rats might shift their decision criteria between tasks. To address this, and to cover the stimulus space more thoroughly, we devised a "mixture identification" task in which we interleaved the full set of stimuli from the categorization and identification tasks as well as intermediate mixtures. Thus, on each trial the stimulus was chosen randomly from one of four mixture ratios at one of four concentrations (Figure 2.5a). Consistent with the previous observations, RTs in this joint task were strongly affected by concentration but not by mixture contrast (Figure 2.5b-c). A two-way ANOVA showed that OSD changed significantly across the different odorant concentrations ($F(3, 48) = 18.57$, $P < 10^{-7}$); but for a given total concentration of the odorants, this change was not significant across the different mixture contrasts ($F(3, 48) = 1.61$, $P = 0.2$). There was no significant interaction of odorant concentration and mixture contrast ($F(9, 48) = 0.20$, $P > 0.9$).

These results indicate that the differences in the relation between accuracy and reaction time for the two tasks are not due to differences in decision criteria. We therefore aimed to unveil the sources limiting performance by taking a computational modeling approach.

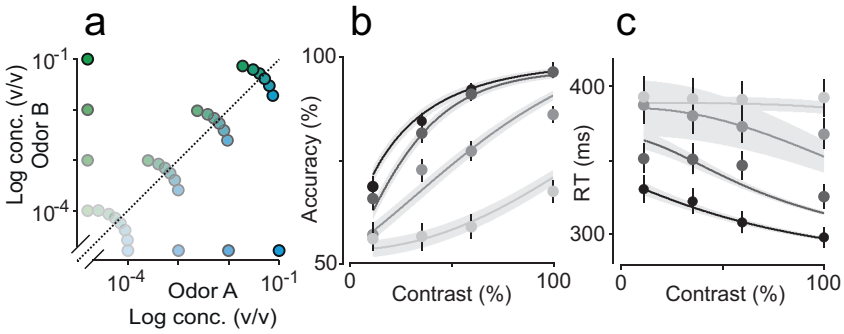

Figure 2.5. Odor mixture identification task. **(a)** Stimulus design. Two odorants (S-(+)-2-octanol and R-(-)-2-octanol) were presented at different concentrations and in different ratios as indicated by dot positions. In each session, four different mixture pairs (i.e. a mixture of specific ratio and concentration and its complementary ratio) were pseudo-randomly selected from the total set of 32 mixture pairs and presented in an interleaved fashion. **(b, c)** Mean accuracy (b) and mean of reaction times (c) plotted as a function of mixture contrast. Each point represents a single mixture ratio. Error bars are mean ± SEM over trials and rats. Solid lines are predictions from the RL-DDM model (Chapter 5), with shades representing the 95% confidence interval simulated for the observed number of trials in the behavioral data. Colors represent the total concentration of the mixture, with black indicating a 10^{-1} mixture and lightest grey 10^{-4} mixtures.

Chapter 3

Noise and uncertainty in olfactory decisions

"The cake is a lie"
– Written in a wall, *Portal 2*

3.1 Chapter Summary

In the previous Chapter we showed that, for a similar level of difficulty, identifying odors at low concentrations requires a much larger increase in stimulus sampling time than does discriminating similar mixtures, even when species, modality and motivation are controlled for.

In this Chapter, by taking a modeling approach, we investigate the role of sensory uncertainty in RTs and response accuracy observed in both identification and categorization task. We also address the possibility of additional sources of noise. This Chapter is divided in 3 sections:

- **Introduction** – a brief introduction regarding sensory uncertainty and accumulator models, in particular the drift-diffusion model.
- **Sensory noise and the diffusion-to-bound model** – in which we explain our model implementation, the data fitting procedures and model comparison methods. We also present the obtained fits for our behavioral data and expose the challenge of fitting both tasks simultaneously.
- **Additional sources of noise** – where we address the DDM's lack of fitting capacity to our data and introduce an additional source of noise, both in the form of a convolution with a Gaussian kernel and random weights fluctuations.

3.2 Introduction

It is commonly assumed that perceptual decisions are difficult because of sensory uncertainty, i.e. noise in the stimulus or stimulus transduction processes (Green & Swets, 1966). Both response time and accuracy depend on the amount of sensory noise (and thus difficulty) of a perceptual judgment. These effects have been long explored in the field of neuroscience (Green & Swets, 1966; Stevens, 1975; Ratcliff, 1978). Theories of stimulus scaling and decision mapping, which are essential for the understanding of accuracy, are the fundamental principles behind signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991). Conversely, response time has long been addressed by sequential sampling theory (Luce, 1986). The crosstalk and interplay between these two fields is what gave rise to accumulation and diffusion models in decision-making (Palmer et al., 2005; Ratcliff & McKoon, 2008).

The phenomenon of stimulus noise is what is believed to be the instigator of SATs in perceptual decision making (Gold & Shadlen, 2007) and can be captured by diverse integration models, such as the accumulator model (Smith & Vickers, 1988), the two-race competition model (Usher & McClelland, 2001) and the drift-diffusion model (DDM; Figure 3.1). These models have been used to explain a wide range of response time and choice behavior data in species from primates (Palmer et al., 2005; Hanks, Kiani, & Shadlen, 2014) to insects (DasGupta, Ferreira, & Miesenböck, 2014).

The DDM posits that decisions are made when a decision variable (DV), whose drift rate μ influences the accumulating evidence, reaches a response bound θ (Ratcliff, 1978; Ratcliff & McKoon, 2008; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Link, 1992). The drift rate is proportional to the strength of the evidence but noise gives rise to variability in the response even upon repeated presentation of the same stimulus (Ratcliff, 1978). The average path of diffusion for a given stimulus

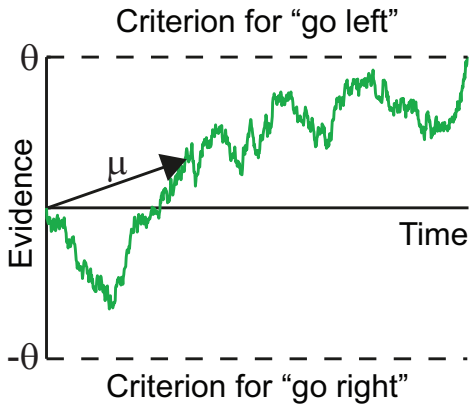

Figure 3.1. Drift-diffusion model. An example path of a “go left” decision. The model integrates noisy momentary evidence over time and makes a decision when one of the criteria is reached. The accuracy and speed of the model depends of the signal-to-noise ratio, i.e. the drift rate μ, here represented as the black arrow.

is what is defined as mean drift-rate. If one considers the variance of the noise process to be constant over time, mean drift rate can be considered a proxy for signal-to-noise ratio (Gold & Shadlen, 2007).

Still, the DDM is not exclusive to sensory uncertainty and noise. Ratcliff and Smith have shown this by adding parameter variability at various levels of the diffusion model (Ratcliff, 1978; Ratcliff & Smith, 2004). These studies focused on differences between correct and error response times and distributions. By looking into a wide range of perceptual and memory tasks, they showed that certain behavioral traits are only explainable if one takes extra forms of variability into account.

Multiple versions of accumulator models have been shown to replicate similar results (Ratcliff & Smith, 2004; Usher & McClelland, 2001; Smith & Vickers, 1988). And in many cases deep-level analysis have revealed that distinguishing between these different models might be an extremely fastidious and difficult task (Ditterich, 2006; Ratcliff & Smith,

2004). However, the DDM does offer the advantage of closed form analytical solutions and a framework that is easy to replicate and test in-silico (Ratcliff, 1978). In fact, given the correct neural-network architecture, the DDM is, in many cases, equivalent to other integration models (Gold & Shadlen, 2007; Ratcliff & Smith, 2004; Churchland et al., 2011; Drugowitsch et al., 2012).

Considering these ideas, we investigated the amount of noise existent in our two tasks, by taking a modeling strategy with the DDM framework as a starting point. However, we decided to implement a flexible model structure (Figure 3.2) that allowed us to change the model iteratively and generalize/hypothesize what might be occurring at a neural level. We did so in order to test which variables are fundamental to explain the differences seen between the identification and categorization tasks and shed light over which sources of uncertainty are needed to explain our behavioral data.

3.3 Sensory noise and the diffusion-to-bound model

In order to explore further which variables might be constraining the rats' behavioral performance, we implemented a diffusion-to-bound model of the kind commonly used to explain performance and reaction time in psychophysical tasks (DDM; Figure 3.2). Perceptual intensity in olfaction, as in other modalities (Palmer et al., 2005; Brunton et al., 2013), can be well-described using a power law (Stevens, 1957, 1975; Wojcik & Sirotin, 2015). We therefore defined the mean strength of sensory evidence μ for each odor using a power law of the odor concentration (Figure 2.4a-b),

$$\mu_i(c_i) = kc_i^{\beta} \tag{3.1}$$

where k and β are free parameters (Palmer et al., 2005). We constrained k and β to be identical between the two odors, which were stereoisomers with identical vapor pressures and similar intensities (Uchida & Mainen, 2003; Taniguchi et al., 1992; Pierce et al., 1995; Laska et al., 2004).

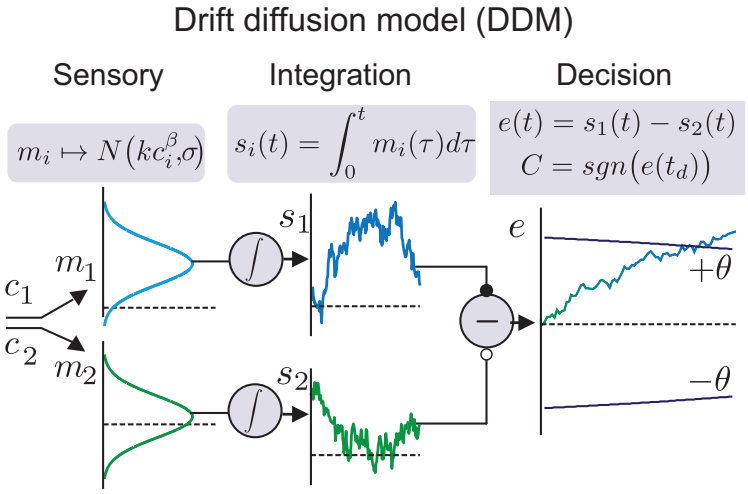

Figure 3.2. Three-layered DDM model. The model consists of three layers Sensory, Integration and Decision. At the Sensory layer, concentrations are transformed into rates that are contaminated with noise σ. These rates are then integrated over time (Integration layer) and combined into a Decision layer. The model's choice will be the bound that is hit first by the evidence particle e. Note that the choice of weights (-1 and 1) for the Decision layer allows it to effectively be a Drift-Diffusion model with collapsing bounds. This model presents 6 parameters.

The DDM is composed of sensory, integration and decision layers. The sensory layer implements a transformation of concentrations into momentary evidence for odors A and B. Evidence at each time step is drawn from a normal distribution $m_i(t) : N(\mu_i, \sigma)$, where $\sigma = 1$ is the standard deviation of the variability corrupting the true rate, μ_i. The integration layer, which also consists of two units, integrates the noisy evidence over time independently for each odor. The last step of the model consists of

a unit that takes the difference between the two integrated inputs. If this difference exceeds a given bound, θ or $-\theta$, the model stops and makes a choice according to the bound that was hit: left for θ, right for $-\theta$. Finally, we allowed for a time-dependent decrease in the bound height ("collapsing bound") with exponential decay rate $1/\tau_\theta$, as has been used to mimic an urgency signal (Churchland et al., 2011; Drugowitsch et al., 2012).

3.3.1 Model implementation

For a given stimulus with concentrations c_A and c_B we define the accumulated evidence at time t, $e(t)$. The diffusion process has the following properties: at time $t = 0$ the accumulated combined evidence is zero, $e(0) = 0$; and the momentary evidence m_i is a random variable that is independent at each time step. We discretized time in steps of 1 ms and run numerical simulations of multiple runs/trials. For each new time step $t = n\Delta t$ we generate a new momentary evidence draw:

$$\mu_i(t) = m_i\big(n\Delta t\big) = N\Big(kc_i^\beta, \sigma\Big) \tag{3.2}$$

that is, through a normally distributed random generator with a mean of kc_i^β, in which we define k as the sensitive parameter, and β as the exponent parameter. σ defines the amount of noise in the generation of momentary evidences. We set σ to 1, making kc_i^β equivalent to the signal-to-noise ratio for a particular stimulus and respective combination of concentrations (c_A, c_B). Integrated evidences (s_1, s_2) are simply the integration of the momentary evidences over time

$$s_i(t) = \int_0^t m_i(\tau) d\tau \tag{3.3}$$

We translate this in our discretized version as a cumulative sum at all time steps, effectively being:

$$s_i\big(n\Delta t\big) = \sum_{j=0}^{n} m_i\big(j\Delta t\big). \tag{3.4}$$

We then define the decision variable accumulated evidence as:

$$e(t) = s_1(t) - s_2(t) \tag{3.5}$$

or in it's discretized version:

$$e\big(n\Delta t\big) = s_1\big(n\Delta t\big) - s_2\big(n\Delta t\big) \tag{3.6}$$

We also define the decision bound and make it collapse over time through an exponential decay. Thus, at timestep n the bound is:

$$\theta(t) = \theta\big(n\Delta t\big) = \theta_0 e^{-n\Delta t/\tau_\theta} \tag{3.7}$$

where we define θ_0 as the bound height at the starting point of integration ($t = 0$) and τ_θ as the bound height mean lifetime. This τ_θ defines the level of urgency in a decision, the smaller it becomes, the more urgent a given decision will become, given rise to more errors (Churchland et al., 2011; Drugowitsch et al., 2012).

The model iteratively compares the accumulated evidence at each time step to the bound, if $e\big(n\Delta t\big) > \theta\big(n\Delta t\big)$ or $e\big(n\Delta t\big) < -\theta\big(n\Delta t\big)$ then the integration process stops. We define time at the decision t_d as the time when crossing occurred, and a choice is made for trial T:

$$C_T = choice = \begin{cases} left, & e(t_d) > 0 \\ right, & e(t_d) < 0 \end{cases} \tag{3.8}$$

and end up setting $e(t_d)$ to $\theta(t_d)$ or $-\theta(t_d)$, according to the crossed threshold. This is important for later versions of the model in order to avoid overshooting learning (Chapter 5).

Additionally our model incorporates a lapse rate for the decision, l_r. That is, for any decision there is a chance l_r that the model will make the opposite decision randomly and regardless of the stimulus being presented. This is necessary to fit a saturated response for the very easy trials, as they never reach complete 100% accuracy. These lapse rates are typically needed for this type of models and have been hypothesized in the past to be due to effects of attention and/or exploration (Wichmann & Hill, 2001).

Lastly, the reaction time for a particular trial is then corrected considering a non-decision time variable t_{ND} that scales the reaction times given by the integration process:

$$t' = t_{ND} + t_d \tag{3.9}$$

The model is run for 1,000,000 trials for each combination of parameters unless when explicitly said otherwise in the text. Free parameters for this particular version of the model are 6: sensitivity (k), exponent (β), non-decision time (t_{ND}), initial bound height (θ_0), collapsing bound mean lifetime (τ_β) and lapse rate (l_r).

3.3.2 Model fitting

We fit our models through log-likelihood maximization (Palmer et al., 2005). However, we differ from that study by using an urgency signal in the form of a collapsing bound, plus trial-by-trial variation of bias and stimulus weights (Chapter 5). For that reason, we cannot apply the same or even any closed form analytical predictions (Ratcliff, 1978). Thus, we simulate numerically every single iteration of the model by discretizing time in 1

ms time bins. For any particular combination of parameters, we simulate 1,000,000 trials. Let $t_T(x)$ be the mean response time of our simulated trials for a given stimulus x. Considering a Gaussian approximation of the reaction time distribution, the likelihood L_T of the observed mean response time $\bar{t_T}(x)$ given the predicted response is:

$$L_T(x) = \frac{1}{\sigma_{\bar{t}}\sqrt{2\pi}} exp\left[- \frac{\big(t_T(x) - \bar{t_T}(x)\big)^2}{2\sigma_{\bar{t}}^2} \right] \tag{3.10}$$

where $\sigma_{\bar{t}}$ is the predicted standard error of the mean given n observed trials. This means that even though the estimate of t_T from the model is obtained by simulating 1,000,000 trials, the expected variance of those responses is set to consider the observed number of trials in the data for a given stimulus. $\sigma_{\bar{t}}$ is also corrected to have a residual time variance of $(0.1t_{ND})^2$ in order to prevent the overall variance from approaching zero in easy trials (Palmer et al., 2005).

For accuracy we assume a binomial distribution. The likelihood of L_P of the observed responses r out of n trials given the predicted simulated proportion P_c is:

$$L_P(x) = \frac{n!}{r!(n-r)!} P_c(x)^r \big(1 - P_c(x)\big)^{n-r} \tag{3.11}$$

The log-likelihoods are then summed over the stimulus strength conditions

$$ln(L) = \sum_x ln\big[L_T(x)\big] + \sum_x ln\big[L_P(x)\big] \tag{3.12}$$

This quantity is maximized iteratively adjusting the model parameters through the usage of Matlab® function *fminsearch* (Nelder-Mead simplex algorithm), by minimizing $-ln(L)$.

3.3.3 Model comparison

Throughout this study we compare the quality of different models in explaining our behavioral data. Direct comparison between models with the same number of parameters can be done by evaluating directly the log-likelihoods. However, this becomes troublesome when comparing models with a different number of parameters.

For comparison between models with different number of parameters we use Bayesian information criterion (BIC) for model selection (Schwarz, 1978). For each model we calculate the BIC (Wit, van den Heuvel, & Romeijn, 2012):

$$BIC = -2ln(L) + q \cdot ln(n) \tag{3.13}$$

where q is the number of free parameters fitted and n the number of data points to be fitted (16 for isolated fits, 32 for simultaneous fits). Each model has a BIC associated to it. The difference between the BICs (ΔBIC) dictates the explanatory strength of one model in relation to the other. The model with the lower BIC is preferred and the evidence in favor is strong if $\Delta BIC > 6$ in comparison with the higher BIC model (Kass & Raftery, 1995). Note that the estimate for BIC in the independent fits (Figure 3.4) is probably under-estimated, as the approximation used for Equation 3.13 considers that $q \ll n$ (Wit et al., 2012).

3.3.4 Fitting individual rats

In order to validate our implemented DDM with collapsing bounds and to control for in-between subject variability, we fitted our DDM to each individual rat and task (Figure 3.3). Each rat presented a clear difference between the amount of RT traded for accuracy dependent on the task at hand and each individual DDM captured this property. By analyzing the DDMs' parameters and their respective values (Table 3.1), we can see that each rat presents a unique solution for each individual task. However, a

closer inspection reveals that for certain parameters this is more prone than for others. Subjective and context dependent parameters such as t_{ND}, integration threshold (Palmer et al., 2005), lapses (Wichmann & Hill, 2001) and collapse bound rate (Drugowitsch et al., 2012) are not surprisingly variable between rats. However, parameters that are related to stimulus such as k and β are more variable between tasks than in between rats, in particular for sensitivity parameter k in which we see a 10-fold decrease for all rats except #2 (Table 3.1). These results indicate that the difference in stimulus between the two tasks plays a more significant effect than in-between subject variability.

Rat	**Task**	k	β	θ_0	τ_θ(ms)	t_{ND}(ms)	l_r(%)	$ln(L)$
#1	Id.	0.325	0.408	8.98	405	252	4.85	-18.4
	Cat.	0.089	0.636	6.37	562	253	0	-21.9
#2	Id.	0.445	0.485	9.52	346	309	1.80	-19.8
	Cat.	0.501	0.861	4.87	547	314	1.66	-21.8
#3	Id.	0.425	0.450	11.7	471	254	2.49	-32.2
	Cat.	0.019	0.513	11.3	775	177	0	-29.9
#4	Id.	0.711	0.514	14.5	552	265	2.60	-40.4
	Cat.	0.013	0.533	15.6	838	107	0	-25.9

Table 3.1. DDM best-fit parameters for individual rats and tasks.

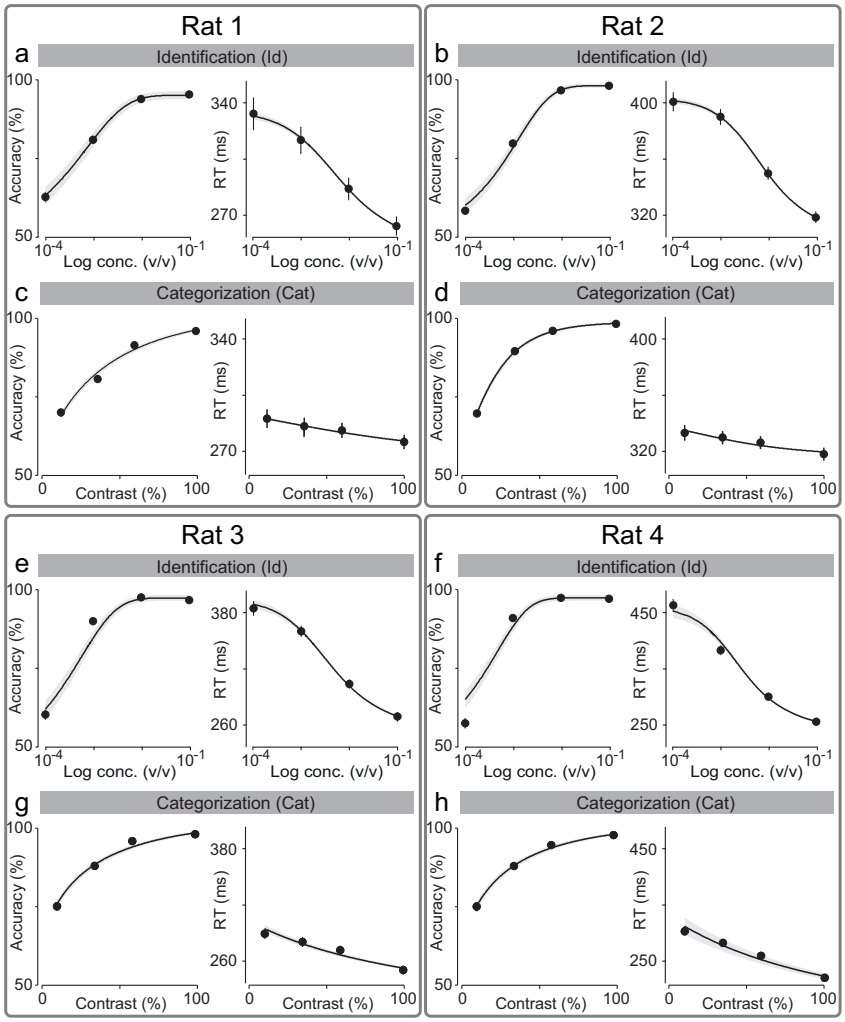

Figure 3.3. Individual behavioral data and DDM fits. **(a,b,e,f)** Behavioral data and fitting results for accuracy and RT in identification task for each individual rat as a function of odor concentration. **(c,d,g,h)** Behavioral data and fitting results for accuracy and RT in categorization task for each individual rat as a function of mixture contrast. Error bars are mean ± SEM over trials. Solid black lines depict each fitted DDM, and shaded area the 95% confidence interval of the model considering the number of trials seen in the behavioral data.

3.3.5 Fitting both tasks simultaneously

As we have seen, the difference in behavior between the two tasks is more significant than individual differences in-between rats (Figure 3.3 and Table 3.1). Thus, we decided to combine all acquired data from the rats and treat our data as a unique dataset. This increased the overall quality of our fits by increasing the number of trials.

We could easily fit both tasks separately and mimic the rats' overall behavior in both tasks (Figure 3.4). In agreement with what was seen individually (Table 3.1; Figure 3.3), the fitted parameters for k were strikingly different (Table 3.2, $k = 1.12$ for identification and $k = 0.08$ for categorization). However, the experiment from the odor mixture identification clearly demonstrated that there must be a convergent solution for the two tasks (Figure 2.5).

Procedure	k	β	θ_0	τ_θ(ms)	t_{ND}(ms)	l_r(%)
Fit Id., pred. Cat.	1.12×10^{0}	0.479	11.1	441	274	3.06
Fit Cat., pred. Id.	7.86×10^{-2}	0.580	10.0	649	220	2.37
Simultaneous fit	2.18×10^{-4}	0.213	14.9	505	180	3.00

Table 3.2. DDM best-fit parameters for identification and categorization tasks.

We thus aimed to find a unique set of parameters that would fit and explain both tasks simultaneously. To do so we developed a series of different fitting procedures in order to test how well this model describes the data. These fitting procedures involved maximizing the log-likelihood function from Equation 3.12 for a data set of 22,126 (identification), 21,371 (categorization) or 43,497 (both) trials using simulations over 1,000,000 trials. The overall quality of the obtained fits with each procedure is shown in Table 3.3.

The first approach we considered was to test whether we could predict the behavioral data of one task using the fitted parameters from the other

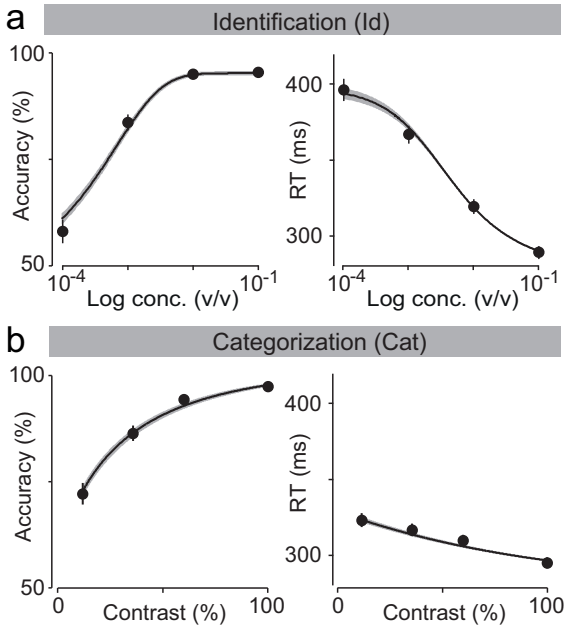

Figure 3.4. Overall behavioral data and DDM fits. **(a,b)** Behavioral data and fitting results for accuracy and reaction time in both identification (a) and categorization (b) task. Error bars are mean ± SEM over trials and rats. Solid black lines depict fitted DDMs, and shaded area the 95% confidence interval of the model considering the number of trials seen in the behavioral data.

task. The free parameters of DDM were first fitted to behavioral data of the identification task. The model captured the overall behavior of the rats in this task. As model performance dropped from 96.9% to 62.9% (data: $97.0 \pm 0.9\%$ to $59.5 \pm 2.7\%$; mean ± SEM, n = 22126 trials/4 rats) with decreasing odor concentration, mean RTs increased from 289 to 393 ms (data: 289 ± 18 to 396 ± 28 ms; Figure 3.5a, black line). This is due to the fact that the evidence for lower concentrations is dominated by noise, making the signal-to-noise ratio smaller. The model therefore takes longer to reach a bound while being more prone to hit the wrong bound first.

Procedure	Number of Parameters	Reference	$ln(L)$ Id.	$ln(L)$ Cat.	BIC
Fit Id., pred. Cat.	6	Fig. 3.5	-47	-1050	2208
Fit Cat., pred. Id.	6	Fig. 3.5	-1759	-47	3628
Simultaneous fit	6	Fig. 3.6	-431	-985	2848
Independent fits	12	Fig. 3.5	-47	-47	220

Table 3.3. DDM goodness-of-fit for identification and categorization tasks.

Next, we asked whether we could predict the rats' psycho- and chronometric curves in the categorization task using the model with the parameters we had fitted to the identification task. As shown in Figure 3.5b, the model had RTs within the same range; model's RT increased from 278 to 306 ms (data: 278 ± 4 to 310 ± 5 ms; mean $\pm$ SEM, n = 21,371 trials/4 rats). But the model strongly overestimated the animals' accuracy at low odor contrast (e.g. model 96.1% vs. data $72.0 \pm 2.5\%$ at 12% mixture contrast; Figure 3.5b, black line).

As a second procedure, we attempted to fit the model parameters to the categorization task and to predict the identification task (Figure 3.5, gray lines). This was also unsuccessful: the model predicted much faster (320 ms) responses than what is seen in the data for lower concentrations.

As we conducted the fits separately the solution might be a compromise between the two tasks. Thus, we changed our fitting procedure of Equation 3.12 to incorporate both identification and categorization data, equally weighted. Nevertheless, this third procedure with simultaneous fits also failed in describing both tasks successfully (Figure 3.6).

The only satisfactory description of our data for this model was to consider both tasks independently (Figure 3.4). However, the experimental data from the mixture identification task (with both sets of stimuli tested in an interleaved fashion; Figure 2.5) ruled out divergent sets of parameters that were achieved in independent fits (Table 3.3). Similar re-

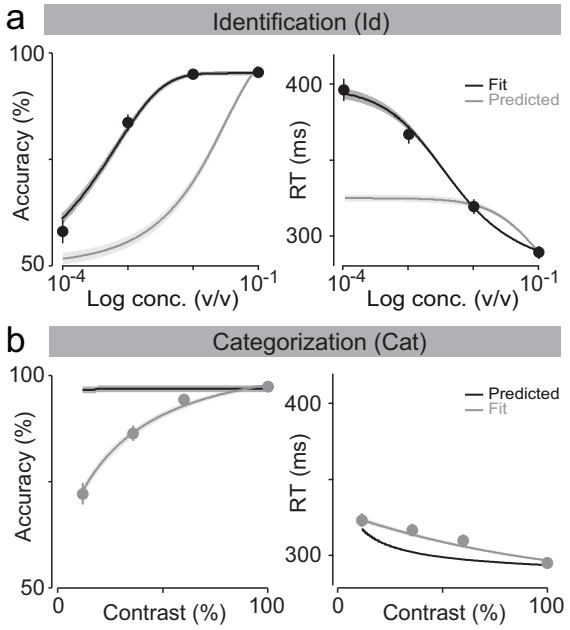

Figure 3.5. Failure to fit performance in one task and predict the other. **(a)** Fitting results for accuracy and RT in identification task. Black solid line represents the model fit for this data, and gray solid lines the prediction from the categorization data fit. **(b)** Fitting results for accuracy and reaction time in categorization task. Solid black lines depict the prediction for this data from the identification fit, and gray solid lines the DDM fit for this data. Error bars are mean ± SEM over trials and rats. Shades for solid lines represent the 95% confidence interval of the model when considering the observed number of trials in the behavioral data.

sults were obtained for other variants of the integration-to-bound model, such as the accumulator model of two-choice discrimination (not shown, Smith & Vickers, 1988) and the two-race competition model (Chapter 5; Usher & McClelland, 2001). These results suggest that this datasets are not compatible with the standard assumptions of this class of models.

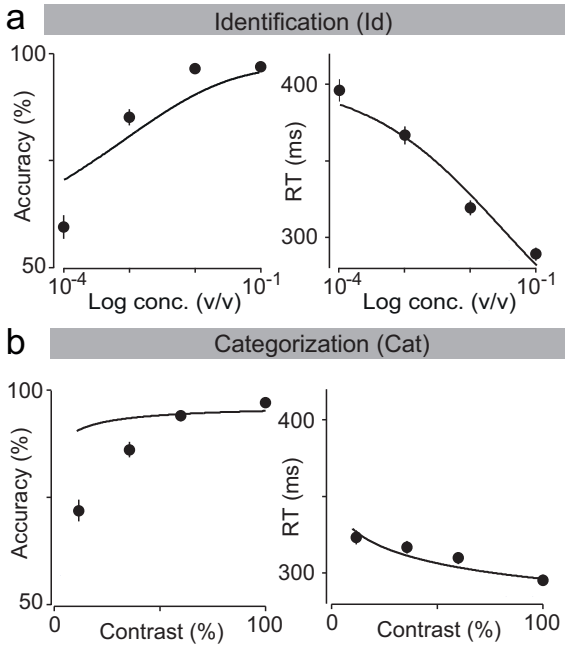

Figure 3.6. Failure to simultaneously fit performance on identification and categorization tasks with DDM. **(a,b)** Fitting results for accuracy and reaction time in both identification (a) and categorization (b) tasks. Black solid line represents the best model fit when considering a likelihood function (Equation 3.12) that takes into account both datasets. Error bars are mean ± SEM over trials and rats.

3.4 Additional sources of noise

Previous work from the lab showed that uncertainty regarding the true category bound plays a key role in the mixture task (Kepecs, Uchida, Zariwala, & Mainen, 2008). In particular, it was shown that rats' decision confidence can be manipulated by varying the distance of the stimulus to the 50/50 separation (Kepecs, Uchida, Zariwala, & Mainen, 2008; Zariwala et al., 2013). Additionally, a computational approach in which a simple discrimination perceptron was contaminated with an extra source

of Gaussian noise (defined as "memory noise") showed aptitude to capture categorization performance (Kepecs, Uchida, & Mainen, 2008).

Our novel identification task allows us to further explore these results by having a better grasp of the amount of sensory noise that the rats are subjected to. In particular, the addition of a DDM allows us to quantify the influence and effects of sensory versus "memory" noise. In this section we explored what occurs when additional noise is added at the decision layer (after integration layer).

3.4.1 Whitening of DDM responses

Our first approach was to consider the simplest version of DDM (Figure 3.1). In this particular case we decided to ignore the collapsing bound so that we could infer the performance and reaction times continuously by taking into account the closed form solutions of DDM (Palmer et al., 2005). These solutions predict that for accuracy the observed curve should be of the form:

$$P_C(x) = \frac{1 - l_r}{1 + e^{-2\theta k|x|^\beta}} \tag{3.14}$$

and for RTs:

$$t_T(x) = \frac{\theta}{kx^\beta} \tanh\left(\theta k x^\beta\right) + t_{ND} \tag{3.15}$$

The introduction of Gaussian noise at the decision layer (Kepecs, Uchida, & Mainen, 2008) can be interpreted in multiple ways: due to an ongoing change of the stimulus representation of the boundary, a noisy memory of that same boundary (Kepecs, Uchida, Zariwala, & Mainen, 2008) or an everlasting change in the rodent's choice function (for example, an ongoing change of the temperature parameter of a "softmax" choice function; Sutton & Barto, 1998). Regardless of the scenario, these hypotheses would generate the same observation at the level of the chrono-

and psycho-metric functions as they would add white noise to the responses of a hypothetically ideal DDM (Ratcliff, 1978).

We tested this by creating a new function for both performance and RTs. We convolve Equations 3.14 and 3.15 to a Gaussian Kernel (Figure 3.7) with zero mean, $g(0, \sigma^2)$.

$$\tilde{P_C}(x) = (P_C * g)(x) = \int_{-\infty}^{\infty} P_C(x - \chi) \exp\Big(- \frac{x^2}{2\sigma^2}\Big) d\chi \tag{3.16}$$

$$\tilde{t_T}(x) = (t_T * g)(x) = \int_{-\infty}^{\infty} t_T(x - \chi) \exp\Big(- \frac{x^2}{2\sigma^2}\Big) d\chi \tag{3.17}$$

which can now be fed into the log-likelihood function from Equation 3.12.

We now force this model to fit both tasks simultaneously, but assuming that all DDM parameters have to be shared between tasks plus σ. The model fits can be found in Figure 3.7 and its respective parameters in Table 3.4.

k	β	θ	$t_{ND}(ms)$	$l_r(\%)$	σ
2.55	0.512	3.54	266	2.86	0.753

Table 3.4. Parameters for DDM convolved with Gaussian kernel.

By adding this "whitening" of the DDM functions we were able to fit both tasks with the same set of parameters. These results indicate that an additional source of variability, that was not originally accounted for in the simple DDM, is influencing performance in these olfactory tasks.

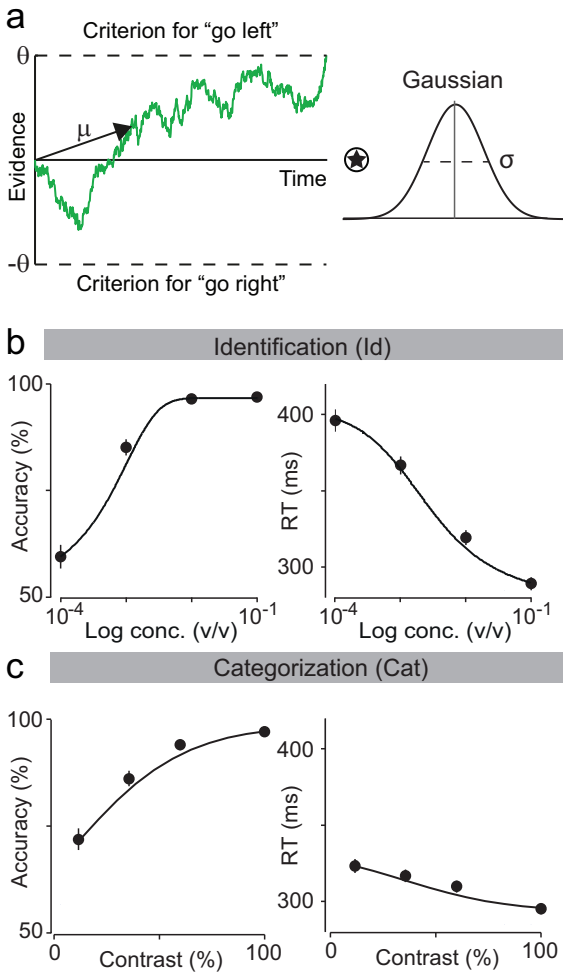

Figure 3.7. DDM convolution with Gaussian kernel. **(a)** Crono- and psychometric DDM functions are convoluted with a Gaussian kernel with width σ. **(b,c)** Fitting results for accuracy and RTs for both identification (a) and categorization (b) tasks. Solid black line – best fitted model to both datasets. Error bars are mean ± SEM over trials and rats.

3.4.2 Random weight noise

We now aimed to understand what is the source behind the whitening of our accumulator model responses. To do that we recovered our original discretized three-layer DDM (Figure 3.2). In that particular version of the model we combined the integrated evidences by considering optimal weights 1 and -1 (Equation 3.5). We now relax this assumption and define w_1 and w_2:

$$e(t) = w_1 s_1(t) + w_2 s_2(t) \tag{3.18}$$

influencing our accumulated evidence particle e on a trial-by-trial basis.

Each trial a new set of weights are picked from two random distribution N_{w_i} with mean 1 and -1 respectively, but with shared variance of σ_w^2 (Figure 3.8). This effectively is equivalent to adding random noise to the drift rates that compose our DDM (Ratcliff & Smith, 2004). This model has the 6 original parameters from the DDM plus the weights noise σ_w. We simulated 1,000,000 trials and fit the model to both datasets iteratively (Figure 3.9). These fluctuations allowed us to find a convergent solution for both tasks (Table 3.5).

k	β	θ	τ_θ	$t_{ND}(ms)$	$l_r(\%)$	σ_w
2.43	0.651	10.1	820	2.86	280	0.409

Table 3.5. Parameters for DDM with random weights noise.

These results show the existence of an additional source of variability that affects the mapping of stimuli into actions. However, the nature of this variability is still unclear as many explanations (or a combination of them) could instigate similar results – random noise injected by the brain, a deliberate strategy from the rats, an imperfect memory representation of the boundary and/or the effects of learning still occurring.

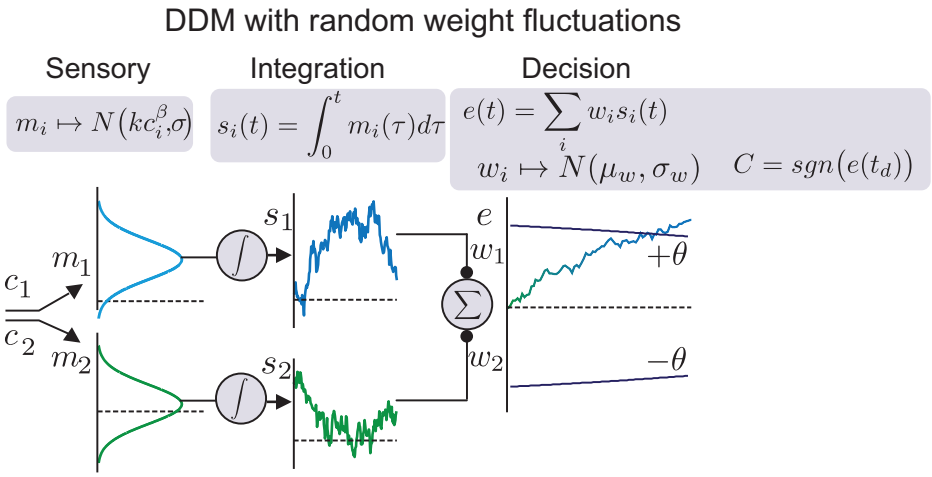

Figure 3.8. DDM with random trial-by-trial weight fluctuations. DDM (Figure 3.2) is expanded to include random weight fluctuation for stimulus weights w_i on a trial-by-trial basis. This is done by feeding the Decision layer with random weights sampled from a Gaussian process. This model presents 7 parameters (6 from the original DDM + 1 more for the variance of the Gaussian process).

In the next Chapters we explore potential sources that could be originating these apparent differences between fast (stimulus related) and slow (trial-by-trial) sources of noise.

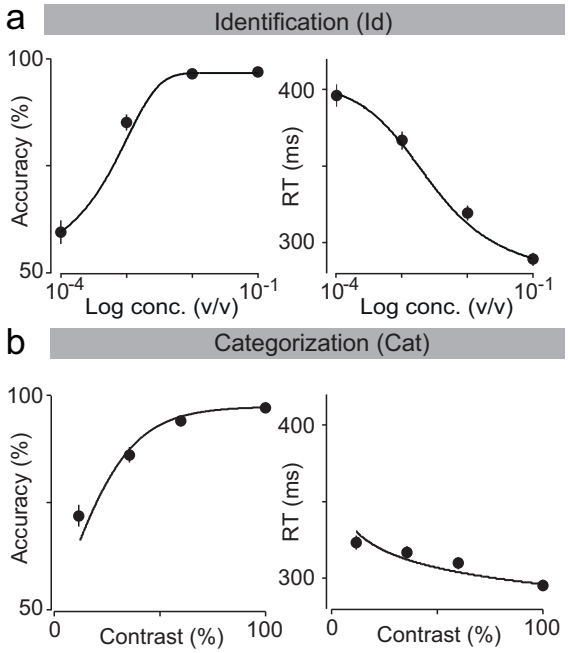

Figure 3.9. DDM with random trial-by-trial weight fluctuations fits both tasks. **(a,b)** Choice accuracy and reaction times in identification task (a) and categorization task (b). Solid black lines show fits of the model to the identification (a), and categorization task (b). Error bars show SEM across rats and trials.

Chapter 4

Reward, errors and biases

"I am error"
– A bearded host, *The Legend of Zelda II*

4.1 Chapter Summary

In the previous Chapter we were unable to fit both odor identification and mixture categorization tasks simultaneously, indicating that stimulus noise was insufficient to explain the observed differences between them. Additionally, we demonstrated that by adding a "slow" type of noise we were able explain both tasks.

In this Chapter, we aim to explore what might be the fundamental principle(s) behind such "slow" noise. We do so by taking a purely behavioral analysis approach of both performance and choice biases. Our analysis is divided at two levels of resolution – session-by-session and trial-by-trial. Thus, this Chapter is divided in three sections:

- **Introduction** – a brief introduction regarding rewards and how Matching law might give rise to changes in choice bias.
- **Session-by-session performance and choice bias** – in this section we look at performance and choice bias fluctuations at a session-by-session level.
- **Trial-by-trial changes in choice bias** – in which we increase our lens resolution and look at performance and bias on a trial-by-trial basis.

4.2 Introduction

UNTIL now we have considered that all uncertainty comes from stochasticity in incoming sensory evidence. However, one possible explanation for the overestimate of accuracy in the categorization task is an additional source of variability that impacts this task preferentially (Uchida et al., 2006; Zariwala et al., 2013). But what is the nature of such variability?

A previous study in rodent visual contrast discrimination showed that, despite obtaining overall high-quality psychometric curves, many sessions presented responses with large biases and high error rates (Busse et al., 2011). The analyzed mice had suboptimal strategies that were influenced by nonvisual factors. In particular, past choices and reward expectations.

The field of decision-making has long recognized reward expectation as playing a significant role in animal behavior (Herrnstein, 1961, 1970, 1974; Herrnstein & Loveland, 1975; Corrado et al., 2005; Kepecs, Uchida, Zariwala, & Mainen, 2008; Zariwala et al., 2013; Lau & Glimcher, 2005). In fact, learning theory focuses on the study of allocation of behavior in response to the rate of received rewards (Sutton & Barto, 1998). Historically, Richard Herrnstein is considered to be the grandfather of reward influences as he proposed the renowned Matching Law (Herrnstein, 1961, 1970, 1974; Herrnstein & Loveland, 1975). ML states that the choice ratio of one particular behavior from a set of behaviors should be proportional to the ratio of the average reinforcement received for that same behavior. This fundamental relationship in neuroscience has now been well established and confirmed across a variety of species (Corrado et al., 2005; Baum, 1983, 1979; Busse et al., 2011; Gallistel, Mark, King, & Latham, 2001; Davison & Baum, 2000)

More recent versions of ML by Baum (Baum, 1974, 1979, 1983), Davison (Davison & Baum, 2000; Baum & Davison, 2004) and Gallistel

(Gallistel et al., 2001; Mark & Gallistel, 1994) have described a wide variety of behaviors under different reinforcement contingencies and reinforcement manipulations (e.g. delay, reward type, value). These observations have brought forward the importance of observing the dynamics of a biological system as it operates in a state of apparent flux.

More specifically and related to the scope of this thesis, matching law and the pursue for rewards have been shown to create biases (Baum, 1974, 1979, 1983). In fact, given the right conditions the emergence of these choice biases might be considered as an optimal strategy (Baum, 1983). Furthermore, reward schedules have been shown to modulate value-based decision making (Corrado et al., 2005; Lau & Glimcher, 2005).

However, reward expectations go to deeper levels than just action selection and value. For instance, reward expectation has been shown to influence performance and RTs in perceptual tasks (Zariwala et al., 2013; Lauwereyns, Watanabe, Coe, & Hikosaka, 2002; Roesch & Olson, 2004), implying that perception can actually be influenced by rewards.

Taking this knowledge into consideration, we hypothesized that ongoing fluctuations in the animals' mapping from odors to choice directions or, equivalently, their representation of the categorical boundary, might limit performance. Such fluctuations might be generated by reward expectations and reinforcement-driven learning processes. To explore this we decided to take a look at performance and choice biases (Busse et al., 2011).

4.3 Session-by-session performance and choice bias

We first verified that the animals' had learned each task fully. To do so we quantified the performance of each rat at a session-by-session level. We consider the number of correct responses r and divide it by the total

number of trials n in a given session j:

$$P_j = \frac{r_j}{n_j} \tag{4.1}$$

We do this for all difficulties, rats, tasks and sessions (Figure 4.1). Despite apparent variability, in particular for the most difficult stimuli, there was no correlation between mean performance and days (red lines in Figure 4.1) for both identification and categorization task in any of the individual rats (Table 4.1).

Rats	**Id.**		**Cat.**	
	ρ	*p-value*	ρ	*p-value*
#1	-0.293	0.13	-0.325	0.24
#2	0.414	0.06	0.089	0.75
#3	0.154	0.62	0.226	0.44
#4	0.055	0.86	-0.240	0.41

Table 4.1. Spearman's rank correlation statistics for session-by-session performance and individual rats.

We next looked at the influence of overall bias in a session-by-session manner. By fitting the derivative of the psychometric curve to a Gaussian distribution we were able to describe the psychometric curve with two parameters, slope (variance) and threshold (mean) (Stevens, 1975; Busse et al., 2011):

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.2}$$

We defined choice bias as the difference between the values of $f(x|\mu,\sigma)$ and chance (0.5) at exactly the middle of the stimulus space ($x = 0$ for the identification task, $x = 50\%$ for the categorization task). We quantified this for each session and individual rat (Figure 4.2). In an analogous way to what was observed for the session performance data (Table 4.1),

statistical analysis on day-by-day session biases showed no correlation over days (Table 4.2).

Rats	**Id.**		**Cat.**	
	ρ	*p-value*	ρ	*p-value*
#1	0.183	0.50	-0.425	0.06
#2	-0.162	0.55	0.264	0.34
#3	-0.379	0.20	0.363	0.08
#4	-0.033	0.92	-0.525	0.06

Table 4.2. Individual rats' Spearman's rank correlation statistics for choice bias on a day-by-day basis.

These results are not surprising when considering the extensive training the rats were subjected to (Figure 2.3). There was no correlation between these performance metrics and session number even at a global scale (considering all rats, Figure 4.3; accuracy: Spearman's rank correlation $\rho = -0.066$, $p = 0.61$ for identification; $\rho = 0.160$, $p = 0.24$ for categorization; bias: $\rho = 0.104$, $p = 0.27$, for both tasks; identification: $\rho = 0.093$, $p = 0.48$; categorization: $\rho = 0.123$, $p = 0.37$).

However, it is clear from observing the overall profiles that there exists unaccounted variability in accuracy and bias across sessions (Figure 4.3).

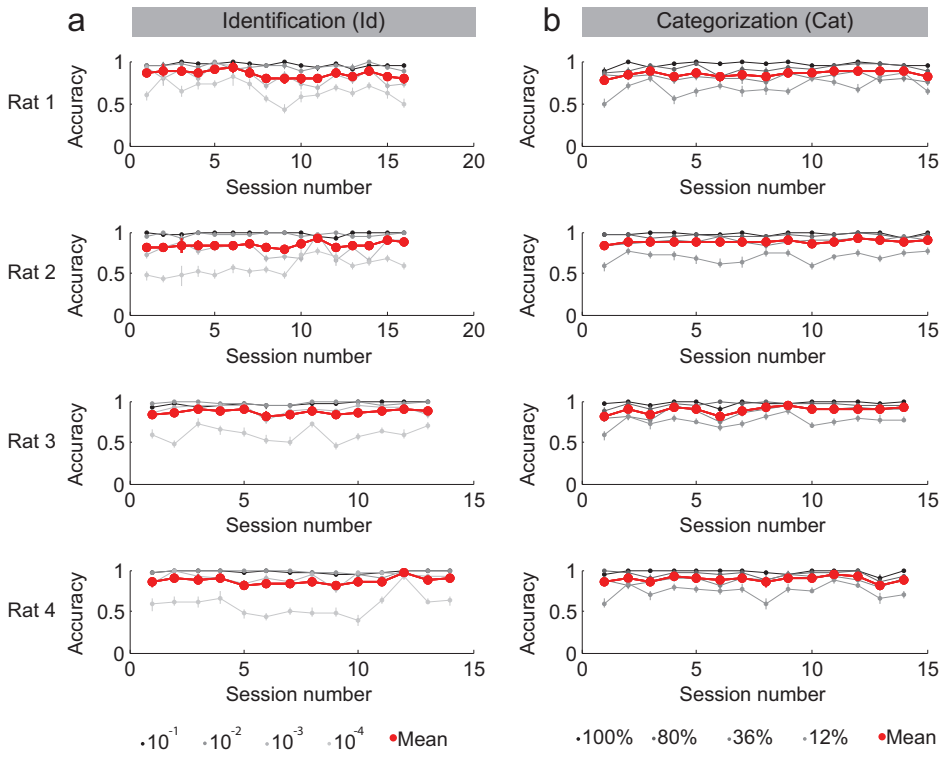

Figure 4.1. Session-by-session stable performance in behavioral data. Performance for each individual rat plotted on a session-by-session basis. Data is presented for both identification (a) and categorization task (b). Depicted are all the sessions used in the analysis. Each difficulty is presented separately. Mean accuracy for all trials is shown in red. Overall performance was stable despite local fluctuations for different difficulties, in particular for most difficult trials. Error bars show proportion SE for each sessions and each individual rat, across trials.

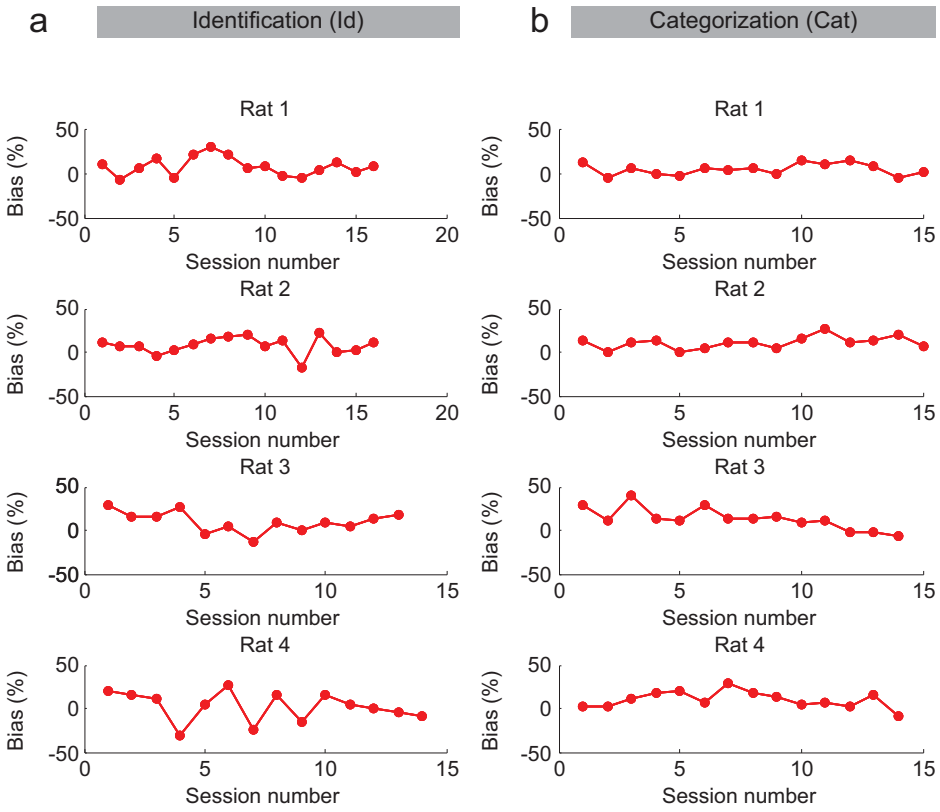

Figure 4.2. Individual rats' session-by-session bias. Session bias (tendency to go left, $B > 0$, or right, $B < 0$) for each individual rat plotted on a session-by-session basis. Data is presented for both identification (a) and categorization task (b). Depicted are all the sessions used in the analysis.

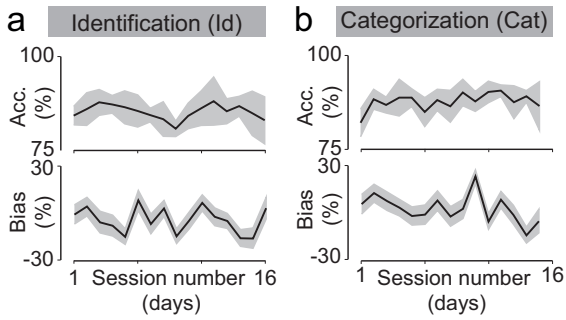

Figure 4.3. Overall session-by-session performance and bias. **(a,b)** Session-by-session change in overall accuracy (top) and performance bias (bottom) for both identification (a) and categorization task (b). Solid black lines are the mean and shaded grey the SEM over the 4 rats for accuracy and 95% c.i. for fitted bias over all rats and trials within a session. Each session was conducted on a different day.

4.4 Trial-by-trial changes in choice bias

In order to explain the potential source of variability seen in choice bias and performance (Figure 4.3) we decided to gain some resolution by exploring the changes at a trial-by-trial level. It has been shown that rodents sometimes exhibit trial-history dependent biases in psychophysical tasks (Busse et al., 2011). However, here we hypothesized that the impact of a past trial would depend not only on the presence or absence of reward, but also on the stimulus that was presented. This would be the case if rats used the stimulus to predict the likelihood of a reward, as dictated by reinforcement learning models (Sutton & Barto, 1998; Rescorla & Wagner, 1972). We therefore examined psychometric curves conditioned on both reward and stimulus (Figure 4.4a).

We did so by conditioning our analyzed psychometric function to trials that followed a correct choice with a given stimulus x_{T-1} (in which $T-1$ represents the preceding trial):

$$f\big(x|\mu_T, \sigma_T, x_{T-1}\big) = \frac{1}{\sigma_T\sqrt{2\pi}} e^{-\frac{(x-\mu_T)^2}{2\sigma_T^2}} \tag{4.3}$$

With this analysis we described conditional psychometric curves that differ depending on the outcome and difficulty of the average preceding trial (Figure 4.4a). We quantified the impact of a previous trial by calculating the difference in the average choice bias conditional upon the trial being correct and a given stimulus being delivered, relative to the overall average choice bias. We considered the original curve (Equation 4.2) and defined the threshold of such psychometric curve as the indifference point ($I = \mu$), which indicates the stimulus difficulty at which the rats have an equal chance of choosing left or right (Figure 4.4a, solid black line). We then quantified the change in choice bias after a given stimulus as the displacement at the indifference point between the two curves:

$$\begin{aligned} \Delta C_B\big(x_{T-1}\big) &= f\big(I|\mu_T, \sigma_T, x_{T-1}\big) - f\big(I|I, \sigma\big) \\ &= f\big(I|\mu_T, \sigma_T, x_{T-1}\big) - 0.5 \end{aligned} \tag{4.4}$$

We calculated $\Delta C_B(x_{T-1})$ (highlighted by the red arrows in Figure 4.4a) for all stimuli and plotted it in Figure 4.4b. Notice the symmetric change is bias, dependent if the rewarded stimulus was following a left choice (positive bias) or a right choice (negative bias). Because $\Delta C_B(x_{T-1})$ was symmetric for left/right stimuli (Figure 4.4b), we plot $\Delta C_B(x_{T-1})$ collapsed over stimuli of equal difficulty (Figure 4.4c). Note that $\Delta C_B(x_{T-1})$ measures the fractional change in choice probability, with $\Delta C_B(x_{T-1}) > 0$ indicating a greater likelihood of repeating a choice

in the same direction as the prior trial and $\Delta C_B(x_{T-1}) < 0$ indicating a greater likelihood of making a choice in the opposite direction.

We also calculated the equivalent updating curves conditional on the trial being incorrect and a given stimulus being delivered (Figure 4.5). These responses are more noisy and thus more difficult to analyze (due to the fact that error trials are less than correct trials). At first glance, these results seem be in agreement with what was reported for rewarded trials, as the sign is in the opposite direction. However, a closer analysis of the curve's shape indicates that the error for that particular trial is more likely to be repeated when compared to the mean response. The explanation for this resides on the fact that our measurement is in relation to the mean psychometric curve, and thus reflecting a bout of incorrect trials. To illustrate this point we plotted the conditional psychometric curves in identification task for trials T and $T-2$ after an easy trial error ($10^{-1}A/0B$) at $T-1$ in relation to the mean psychometric curve (Figure 4.6). What we saw was a change in the direction of the psychometric curve towards the mean (from green curve towards red in Figure 4.6), as expected from prediction error, albeit it not being sufficiently large to move above the mean performance.

In general, rats showed a tendency to repeat a choice in the same direction that was rewarded in the previous trial ("win-stay"; Figure 4.7), consistent with previous results (Busse et al., 2011). But this analysis revealed a qualitative difference between the identification and categorization tasks with respect to how the stimulus in the past trial impacted the change in bias (Figure 4.4). For the identification task, the influence of the previous trial was stimulus independent (Figure 4.4e, one-way ANOVA, $F(3, 12) = 2.0$, $p = 0.17$). In contrast, for the categorization task (Figure 4.4f) the influence of the previous trial showed a graded dependence on the stimulus, being larger for a difficult previous choice than for an easier one ($F(3, 12) = 25.4$, $p < 10^{-5}$). For error conditioned updating curves it

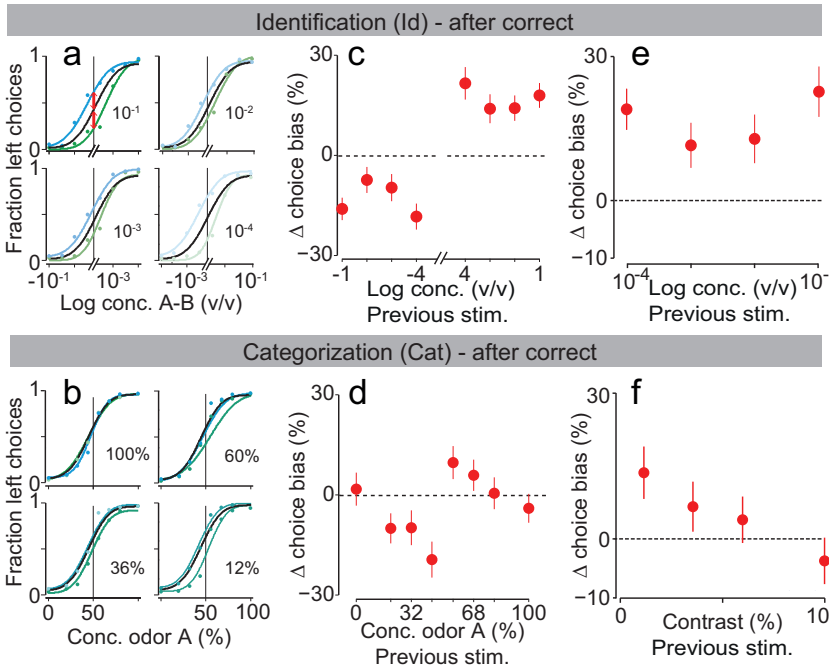

Figure 4.4. Reward history effects are still present despite overall stabilized performance. **(a,b)** Psychometric functions for mean (black solid line) and conditional curves following a reward and a given stimulus. For the identification task all four stimuli from 10^{-1} to 10^{-4} are depicted (a). For the categorization task from 100 to 12% mixture contrast (b). Green points and lines represent trials following a right-reward ($B > A$) and blue trials following a left-reward ($A > B$). Lines represent fits of a cumulative Gaussian to the data (Busse et al., 2011). Red arrows in (a) elucidate the measured change in choice bias. **(c,d)** Full updating curves for change in choice bias. Change in choice bias after a correct choice and a given stimulus. Data presented for identification (c) and categorization task (d). **(e,f)** Collapsed change in choice bias plotted as a function of the previous difficulty, for plots in (a,b). All four different odor concentrations for identification task (e); and all mixture contrasts for categorization (f). Error bars represent SEM over 4 rats.

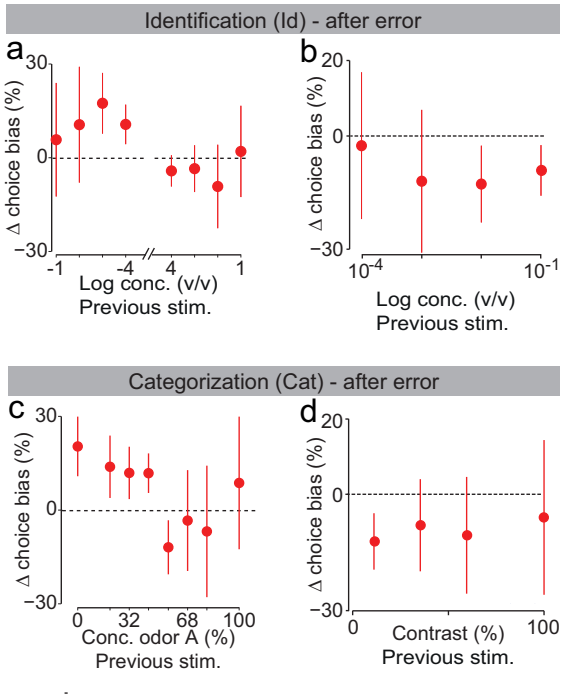

Figure 4.5. Updating curves for change in choice bias after an error response for both tasks. **(a,b,c,d)** Change in choice bias after an error and a given stimulus. Data presented for identification (a,b) and categorization task (c,d). (a,c) Change in choice bias modulated by stimulus concentration. (b,d) Change in choice bias modulated by stimulus difficulty (contrast). Curves show a symmetric effect to what is seen in correct trials. Error bars show SEM across rats.

was not possible to infer with reliability the stimulus dependence due to the smaller number of trials (Figure 4.5).

These results were also consistent at the individual rat level, as all rats show no effect of stimulus difficulty in change of choice bias for the identification task (rat 1, $F(3, 59) = 2.5$, $p = 0.069$; rat 2, $F(3, 59) = 2.34$, $p = 0.08$; rat 3, $F(3, 55) = 2.46$, $p = 0.07$; rat 4, $F(3, 51) = 1.54$, $p = 0.21$). On the other hand, for the categorization task all rats showed stimulus

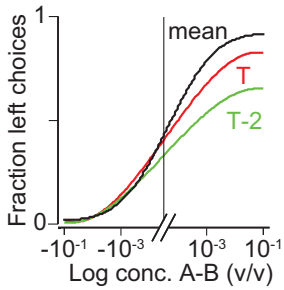

Figure 4.6. Conditional psychometric curves for trials T and $T-2$ after an easy trial error in identification task. Mean psychometric curve for all trials is shown in black. Green line depicts the psychometric curve for all trials that preceded an easy left decision error; red line the psychometric curve in trials that followed that same error. Note that performance improved in relation to the green line, despite the red line being below the black line. This fact explains the inversion seen in Figure 4.5.

difficulty modulation in choice biases (rat 1, $F(3,55) = 4.39$, $p < 0.01$; rat 2, $F(3,55) = 11.7$, $p < 10^{-5}$; rat 3, $F(3,55) = 22.7$, $p < 10^{-4}$; rat 4, $F(3,51) = 3.35$, $p < 0.05$).

These results suggest that rewards have an important role in these two tasks. However, the effects of reward expectations were unaccounted for in our previous DDM. More specifically, the interaction with stimulus difficulty seen for the categorization task led us to hypothesize that reinforcement learning plays a vital role, despite the fact that at a higher resolution the rats seem to not be improving over sessions (Figure 4.3).

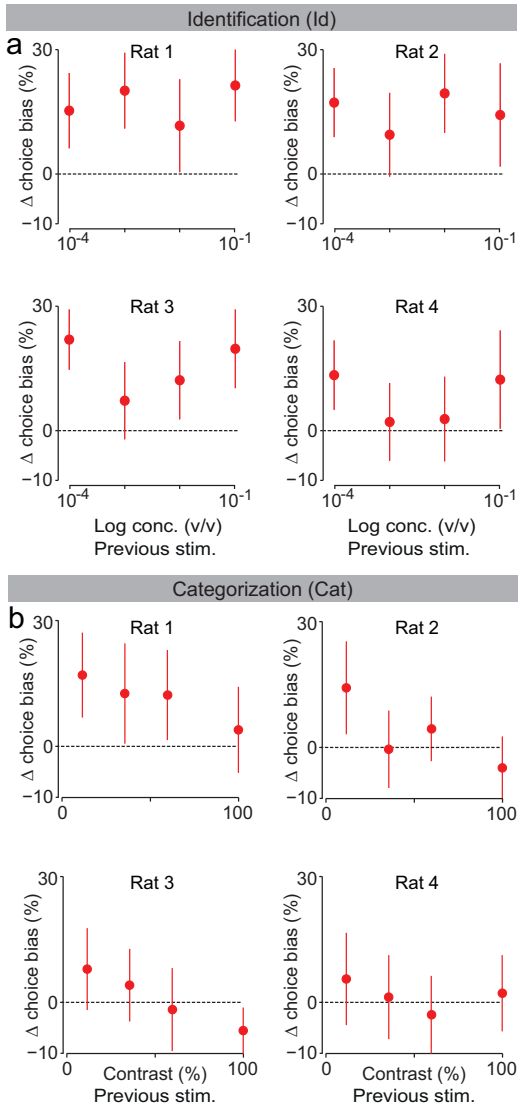

Figure 4.7. Updating curves for change in choice bias after rewarded trials for all individual rats and tasks. **(a,b)** Change in choice bias after rewarded choice and a given stimulus. Data presented for identification (a) and categorization task (b) for all 4 rats. Error bars show SEM across sessions.

Chapter 5

Reinforcement learning and uncertainty

"Everything not saved will be lost"

– Quit screen message, *Nintendo*

5.1 Chapter Summary

In the previous Chapter we saw that despite the fact that performance was not improving over sessions, there was an effect of "updating" in the rats' choice biases on a trial-by-trial basis. We further demonstrated that the nature of this updating was task specific.

In this Chapter, we aim to explore what might be the rules behind these updating effects. We do so by combining our DDM model with an RL delta rule. We will see that our model is able to fit both tasks simultaneously and predict a series of different behavioral results, validating our proposed framework. This chapter is divided in four sections:

- **Introduction** – a brief introduction about signal detection theory, sequential analysis and reinforcement learning.
- **DDM with side bias** – in this section we implement a model that includes a basic form of side bias that follows rewards and avoids errors. We demonstrate that just reward matching is insufficient to explain our behavioral data.
- **Reinforcement learning with DDM - the adaptive diffusion-to-bound model** – where we further extend our model to now include stimulus dependent learning, fit the behavioral data and generate predictions. These predictions are then compared to the behavioral data. This section is the fundamental core of this thesis as it exposes its main findings.
- **The two-race accumulator model** – in which we present a second version of an accumulator model combined with a RL rule. We demonstrate how these type of models can be generalized to different forms.

5.2 Introduction

We have seen that there is a difference in the way speed and accuracy trade in odor identification and mixture categorization. In regards to the identification task the amount of SAT is much larger than categorization. That difference is probably due to the fact that the most difficult stimuli in the identification task are below detection threshold and thus signal-to-noise ratio massively reduced.

In signal detection theory (SDT) the decision-maker obtains an observation of noisy evidence from the stimulus, that observation is then evaluated according to a given decision rule (Green & Swets, 1966; Stüttgen, 2011). In simple binary decisions, the decision variable is typically related to the likelihood ratio of the different alternatives, and then compared to a criterion (Gold & Shadlen, 2007; Glimcher & Fehr, 2014). Sequential analysis (SA) is a natural extension to SDT that combines multiple pieces of evidence observed over time. Conceptually, the idea is that in the presence of uncertainty or noise, the decision maker can benefit from sampling multiple times from the noisy distribution of values representing the stimulus (Luce, 1986; Townsend & Ashby, 1983; Vickers, 1970; Vickers et al., 2014; Gregson, 2014; Usher & McClelland, 2001; Ratcliff & Smith, 2004). The parameters of these sequential sampling models allow quantifying several latent psychophysical processes, namely the speed of sensory information processing, given by the rate of accumulation; response caution, from the bound height; and the amount of time spent on processes unrelated to decision formation (Ratcliff & McKoon, 2008). In recent years, integrator models have gained prominence in the field of decision-making due to their ability to explain several features of behavioral and neural data (Ratcliff, Van Zandt, & McKoon, 1999; Palmer et al., 2005).

However, decisions might be exposed to noise at a second layer that is not addressed by SDT or the SA framework in its simplistic form (although

see Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011; Krajbich, Lu, Camerer, & Rangel, 2012; Milosavljevic, Malmaud, Huth, Koch, & Rangel, 2010 for applications of integration models in value-based choices). In fact, our lack of fitting capacity of a simple DDM that has shared properties between the two tasks points in that direction. Specifically, we saw that there is a "slow" component of noise that is not integrated within the course of a trial but shows influence on a trial-by-trial basis.

This has not been the first time these influences have been observed in rodents (Busse et al., 2011). In that particular study, Busse et al. combined a SDT framework with a value-based decision model (Corrado et al., 2005; Lau & Glimcher, 2005) in order to create an extended probabilistic choice model. This model was successful in explaining the observed effects of choice bias in a 2AFC task. However, this model is related to a general bias that might be reflecting a form of Herrnstein's matching law (Herrnstein, 1961, 1970, 1974; Herrnstein & Loveland, 1975; Baum, 1983; Busse et al., 2011). For our particular case the existence of a stimulus dependent change in bias appears to be a missing piece of the puzzle.

In this Chapter we focused our attention in reinforcement learning (RL). In classical conditioning, reinforcers are delivered independently of any action taken by the animal (Rescorla, 1988). In instrumental conditioning, the actions of a subject determine what reinforcer is obtained (Dickinson, 1980; Mackintosh, 1983; Shanks, 1995). Learning about stimuli or actions taking into account just rewards and errors (punishments) is the core concept behind the principles of RL (Sutton & Barto, 1998, 1990). In both classical and instrumental conditioning predicting a reward has been shown to play an important role in explaining a multitude of different behaviors (Shanks, 1995; Sutton & Barto, 1990). In particular, the Rescorla-Wagner delta-rule (Rescorla & Wagner, 1972) gives a concise account of certain aspects of conditioning and behavior (Mackintosh, 1983; Gallistel, 1993; Shanks, 1995).

What we aimed to do in this Chapter was to combine these two fields of knowledge. We hypothesized that our two tasks exist at a plane in which the interaction between a detection process and a value-guided choice is of paramount importance.

5.3 DDM with side bias

The most simplistic version of a standard diffusion-to-bound model does not predict changes in choice bias at any level (both at session-by-session and trial-by-trial levels). We therefore sought to extend the model with an extra source of variability that can account for the choice biases. Our first approach was to consider a general reward bias that has been reported in rodents performing a visual discrimination task (Busse et al., 2011). Our first question was to address whether this model could replicate what was seen in the data and in particular for the categorization task.

5.3.1 Model implementation

This first extended version of our three layered DDM (Figure 5.1) is in all ways similar to the previous one (Figure 3.2), except for two parts: when defining the accumulated evidence e; and a trial-to-trial dependency that was not present in the previous model.

We define bias term b, a scalar that is taken into account at each trial. This scalar sets the initial value of accumulated evidence:

$$e(t = 0) = e_0 = w_b b = w_b \tag{5.1}$$

w_b is a weight parameter that determines how strong b is for this particular trial. As b is simply a constant that is modulated by the value of w_b (i.e. they tradeoff) we set it to 1. The process then continues normally with

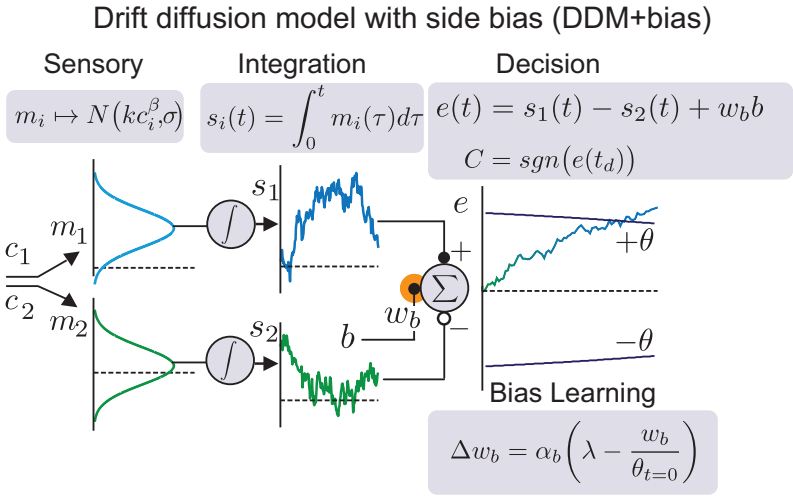

Figure 5.1. Drift-diffusion model (Figure 3.1) with reward bias. DDM is expanded to include reward side dependencies. This is done by feeding the Decision layer with a constant bias ($b = 1$). The magnitude of this effect is set by updating the weight w_b each trial, after the model accesses the choice outcome. Note that this effectively means a change in the DDM's starting position on a trial-by-trial basis. This model presents 6 parameters.

the accumulated evidence being:

$$e(n\Delta t) = s_1(n\Delta t) - s_2(n\Delta t) + e_0 \tag{5.2}$$

Decision bound is crossed and the decision set as in the previous model.

To reflect a simple RL process, this w_b is updated according to a delta rule (Widrow & Hoff, 1960). After the decision and before the start of the next trial, this updating occurs adaptively and is designed to follow an unsupervised reinforcer that depends on side choice,

$$\Delta w_b = \alpha_b\left(\lambda - \frac{e_0}{\theta_0}\right) = \alpha_b\left(\lambda - \frac{w_b}{\theta_0}\right) \tag{5.3}$$

with λ (the rewarded outcome of the trial) equal to $+1$ if the rewarded choice was to the "left", -1 if it was "right" and 0 if the trial was unrewarded, and α_b the learning rate for the bias.

An unbiased and symmetric DDM will show a probability of 50% of reaching any bound when the drift rate is equal to zero (Palmer et al., 2005). The introduction of this parameter is equivalent to changing the starting position of the diffusion process and thus making the integration asymmetric (Ratcliff & McKoon, 2008; Hanks et al., 2014). Therefore, the 50% unbiased probability is now incorrect. Let us consider the upper bound as A and the lower bound as $-B$. This implies that the probability of reaching bound A with no stimulus input is (Palmer et al., 2005):

$$\lim_{\mu \to 0} P_a = \frac{B}{A+B} \tag{5.4}$$

In our particular case the starting position is changed by w_b. Analogously we can then write $A = \theta_0 - w_b$ and $B = \theta_0 + w_b$. Meaning that we can re-write Equation 5.4 as:

$$\lim_{\mu \to 0} P_a = \frac{\theta_0 + w_b}{\theta_0 - w_b + \theta_0 + w_b} = \frac{1}{2} + \frac{w_b}{2\theta_0} \tag{5.5}$$

In an unbiased DDM $\lim_{\mu \to 0} P_a = 1/2$. We assume that in average, for all trials, this is probably the case. However, to mimic the effects seen in Figure 4.4e we define the average change bias we observe in the data as:

$$\big\langle \Delta C_B(x|r=1) \big\rangle = \frac{|\Delta C_B(x|\lambda=1)| + |\Delta C_B(x|\lambda=-1)|}{2} \tag{5.6}$$

that is, conditioned on previous choice and outcome. We assume that our model will mimic closely this change in choice bias with a weight change of the size of:

$$\Delta w_b = \frac{\alpha_b}{2\theta_0} \approx \big\langle \Delta C_B(x|r=1) \big\rangle \tag{5.7}$$

and thus constrain the bias learning rate to be:

$$\alpha_b = 2\theta_0 \big\langle \Delta C_B(x|r=1) \big\rangle \tag{5.8}$$

Initial value for w_b is set to zero (unbiased). The model is run for 1,000,000 trials for each combination of parameters. Free parameters for this particular version of the model are 6, sensitivity (k), exponent (β), non-decision time (t_{ND}), initial bound height (θ_0), collapsing bound mean lifetime (τ) and lapse rate (l_r).

As described above, both tasks could be fit independently using the simple DDM (Figure 3.4) but could not be fit simultaneously (Figure 3.5 and 3.6). In order to understand the influence of w_b in the two tasks we decided to focus our approach: first fitting one task (identification) and then attempting to predict the second task with the same set of parameters (categorization).

5.3.2 Model fitting results

The obtained fits for DDM with reward bias (DDM+bias) can be found in Figure 5.2 and Table 5.1. We observed that DDM+bias fits the accuracy and reaction time for the identification task at all difficulty levels (Figure 5.2a,b). It also captured the correct sign and relative magnitude of the history-dependent changes in choice bias observed in the identification tasks (Figure 5.2c). However, it still failed to predict the chronometric and psychometric curves of the categorization task. The model strongly outperformed the rats (Figure 5.2d) while the reaction time changes were roughly comparable (Figure 5.2e). More importantly the model failed to reproduce the stimulus dependency seen for the updating curves in categorization or any effect for that matter (Figure 5.2f).

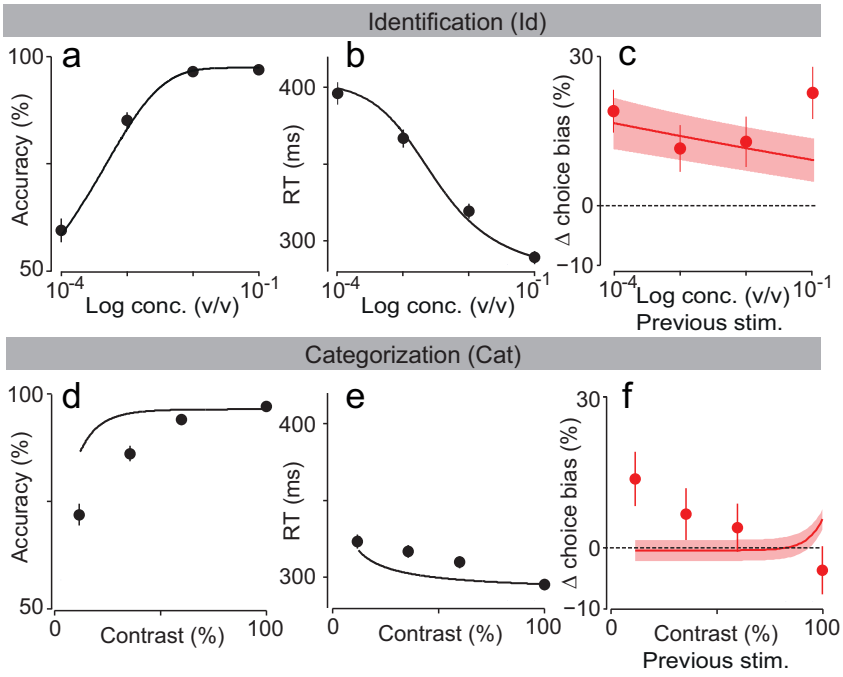

Figure 5.2. DDM+bias fails to predict categorization task. **(a,b,c)** Model fits to identification task data. Choice accuracy (a), reaction times (b) and updating curves (c) for identification task. **(d,e,f)** Model predictions for categorization task. Choice accuracy (d), reaction times (e) and updating curves (f) for categorization task. Error bars show SEM across rats and trials. Solid lines show fits (a,b,c) and predictions (d,e,f) from the DDM+Bias model, with shaded areas representing the 95% c. i. expected by considering the same number of trials as the data.

5.4 Reinforcement learning with DDM - the adaptive diffusion-to-bound model

The lack of updating in the categorization task for DDM+bias (Figure 5.2f) led us to implement an additional model in which stimulus learning must be present and play an important role. The form of the updating

curve for the categorization task appears to be consistent with the predictions of an RL process that is driven by a quantitative prediction error. For correct trials, rewards are more surprising (i.e. larger prediction error) when they follow difficult stimuli than when they follow easy stimuli, because difficulty sets the level of reward prediction (Kepecs, Uchida, Zariwala, & Mainen, 2008). It is less clear why the identification task would not also be also be associated with similar updating curves. We reasoned that the identification task would be more subjected to stimulus-independent choice biases (Busse et al., 2011) since difficult identification stimuli reflect low stimulus intensity, whereas difficult mixture stimuli do not.

To test these ideas, we implemented a diffusion-to-bound model with stimulus dependent RL (which we refer to as "RL-DDM", Figure 5.3).

5.4.1 Model implementation

We implemented RL-DDM by augmenting the DDM+bias allowing not only a change of starting position, e_0, but also introducing two new weights that scale the quality of the integrated information for each odor, w_1 and w_2. These weights transform each stimulus input into evidence:

$$e_i = w_i s_i(t) \tag{5.9}$$

which is then combined to form a net evidence

$$e(t) = w_1 s_1(t) + w_2 s_2(t) + w_b b. \tag{5.10}$$

or in the discrete formulation:

$$e(n\Delta t) = w_1 s_1\big(n\Delta t\big) + w_2 s_2\big(n\Delta t\big) + e_0 \tag{5.11}$$

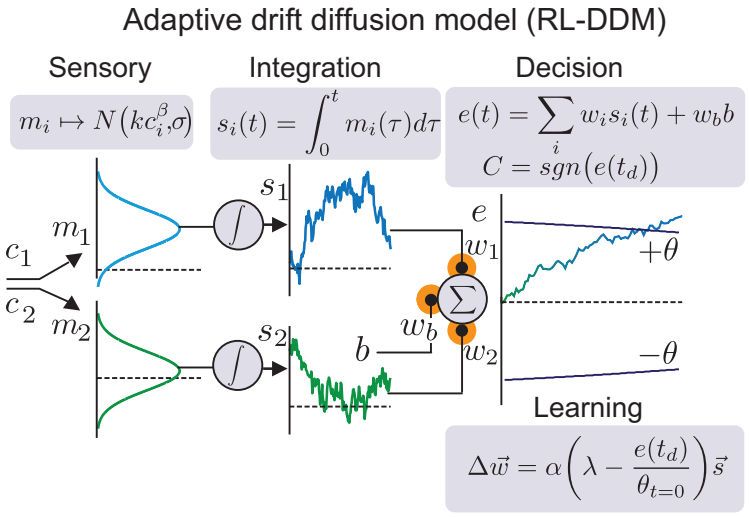

Figure 5.3. Drift-diffusion model with bias and stimulus learning. Integration DDM (Figure 5.1) is even further expanded with the addition of changing stimulus weights, w_1 and w_2. These weights are then combined with the integrated momentary evidences (s_1,s_2) plus the offset set by the bias weight w_b. After each trial the model updates stimulus weights according to the obtained outcome through a delta-learning rule. This model has 7 parameters.

In this model there are two types of weights: w_b that reflects the strength of stimulus-independent bias and w_i which provide multiplicative scaling of the sensory evidence input of the integrator.

Whereas in the standard DDM, stimuli were assumed to be symmetric and opposite, i.e. $w_1 = -w_2$, in the RL-DDM, this constraint was relaxed. The weights, along with the input bias term, defined the location of the category "boundary" in stimulus space, a line that divides stimuli associated with right and left choices (see Figure 2.2). These weights were fixed within a trial and updated between trials, thereby adding a source of trial-by-trial variability to the model. These fluctuations are equivalent to

variability in the drift rate (Ratcliff, 1978), though not from a stochastic source, but from a deterministic learning rule.

We updated the stimulus weights with a variation of the delta rule. But, critically, the error signal was based on the difference between the actual outcome and a quantitative estimate of the expected outcome. To achieve this, we used the difference between the rewarded outcome of the choice (λ, taking the values -1, 1 and 0, as before) and the relative amount of sensory evidence at the time of the choice, $e(t_d)$. We normalized by dividing the maximum possible evidence, which is the value of the bound at time 0, θ_0, giving:

$$\Delta\vec{w} = \alpha\left(\lambda - \frac{e(t_d)}{\theta_0}\right)\vec{s} \tag{5.12}$$

in which α is the learning rate. After updating, the weights are normalized before starting the next trial through the following rule (Dayan & Abbott, 2005):

$$\vec{w} = \frac{\vec{w}}{\parallel \vec{w} \parallel} = \frac{\vec{w}}{\sqrt{w_1^2 + w_2^2}} \tag{5.13}$$

This normalization step is not necessary, but important to avoid explosion in weight values and thus stabilize the fitting procedure. All the results presented in this thesis have been validated with this normalization, with a different type of normalization (cutoff at 1 and -1) and without (data not shown).

Initial value for w_b is set to zero (unbiased), and (w_1, w_2) to perfect drift-diffusion ($1/\sqrt{2}$, $-1/\sqrt{2}$). The model is run for 1,000,000 trials for each combination of parameters unless when explicitly detailed otherwise. Free parameters for this version of the model are 7 – sensitivity (k), exponent (β), non-decision time (t_{ND}), initial bound height (θ_0), collapsing bound mean lifetime (τ), lapse rate (l_r) and stimulus learning rate (α).

This learning rule introduces just one new parameter, the learning rate, α. To see the effect of this rule, consider that the stopping rule dictates that sensory evidence $e(t_d)$ must reach the bound for a decision to be taken. However, because the bound height is collapsing, i.e. monotonically decreasing in time according to $\theta(t) = \theta_0 e^{-t/\tau}$, the evidence at the time the choice is made must decrease in time. Trials with more difficult stimuli and longer decision times will therefore be associated with on average lower evidence at decision time (Churchland, Kiani, & Shadlen, 2008; Drugowitsch et al., 2012); this implies more weight change or updating.

The bias term, w_b, represents stimulus-independent effects such as biases associated with the reward port itself. It is also equivalent to specifying the starting position in the drift-diffusion process or changing the bound heights (Ratcliff, 1978; Mulder et al., 2012). As stated above, the input b is constant and unvarying from trial to trial ($b = 1$), but w_b is updated on a trial-by-trial manner though independently of the presented stimulus through Equation 5.3 with learning bias set by Equation 5.8. The amount of change in w_b is set to mimic the stimulus independent change in bias component seen in the data (Figure 5.2c).

Note that by adding trial-by-trial learning we have effectively created a model that not only has integration of evidence within a trial, as it also integrates across recent trial outcomes. Furthermore, by separating the stimulus dependent and stimulus-independent components we can distinguish the effect of reward on bias and its interaction with stimulus.

5.4.2 Model fitting

Here, to test the ability of the RL-DDM to capture performance in both tasks, we again used the procedure of first fitting one task (identification) and then attempting to predict the second task (categorization). The addition of the reinforcement learning rule preserved the ability of the model to fit the identification task (Figure 5.4). This is expected, as

the DDM was already doing a good job fitting this task (Figure 3.5). Visual inspection of the fitting landscape showed that the RL-DDM was not well-constrained by the identification data (Figure 5.5a). We therefore attempted to fit the odor categorization data by only changing the learning rate, α. We implemented this by changing slightly our original fitting procedure.

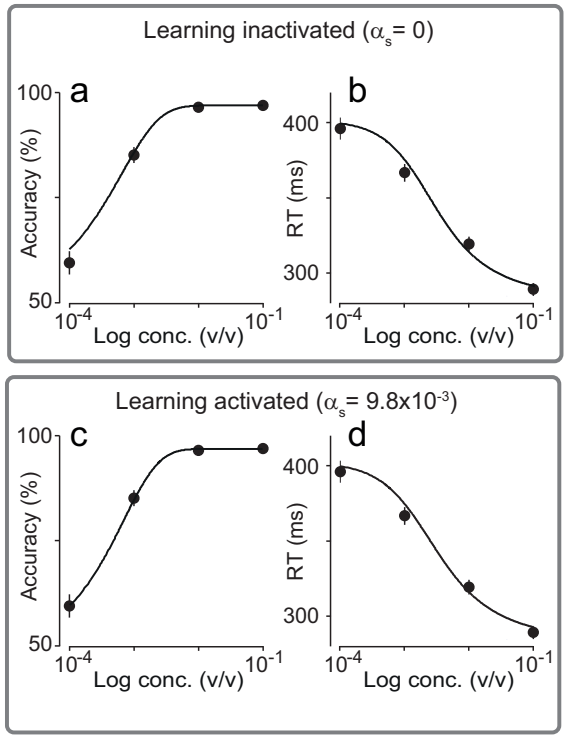

Figure 5.4. Stimulus learning influence on identification task. **(a,b)** Accuracy and reaction times plotted as a function of odor concentration (v/v) for the identification task. Line indicates DDM fit from Figure 3.4a. **(c, d)** Same as (a,b) but with stimulus learning activated with the best estimated parameter to fit categorization task.

We expanded the likelihood function (Equation 3.12) to constrain the learning parameter. For the identification data we allow all the parameters to change except the stimulus-learning α.

$$\ln(L_I|\alpha) = \sum_c \ln\left[L_T(c|\alpha)\right] + \sum_c \ln\left[L_P(c|\alpha)\right] \tag{5.14}$$

Where c represents each individual concentration stimulus given. For the categorization data we allow α to change given that all the other parameters are held fixed.

$$\begin{aligned}\ln\left(L_C|k,\beta,t_{ND},\theta_0,\tau,l_r\right) = &\sum_m \ln\left[L_T\big(m|k,\beta,t_{ND},\theta_0,\tau,l_r\big)\right] \\ &+ \sum_m \ln\left[L_P\big(m|k,\beta,t_{ND},\theta_0,\tau,l_r\big)\right]\end{aligned} \tag{5.15}$$

Where m represents each individual mixture. We then maximize the sum of the two log-likelihoods:

$$\ln\big(L\big) = \ln\big(L_I|\alpha\big) + \ln\big(L_C|k,\beta,t_{ND},\theta_0,\tau,l_r\big) \tag{5.16}$$

We found that it was now possible to fit simultaneously the psychometric and chronometric curves of both identification and categorization tasks with a single set of parameters. Note that the inclusion of α had a reduced effect for identification when the best fit parameter α was activated (small decrease of performance for 10^{-4}, however this was not a significant effect, t-test $p = 0.75$). This justified the inclusion of the fitting rule from Equation 5.16. Visual inspection of the fitting landscapes showed that, by taking into account the categorization data and Equation 5.16, α is now better constrained (Figure 5.5b).

Lastly, we tested that our fitting procedure was converging to a stable solution. Visual inspection of the log-likelihood fitting landscapes showed that our model is converging into the maximum value possible (Figure 5.6,

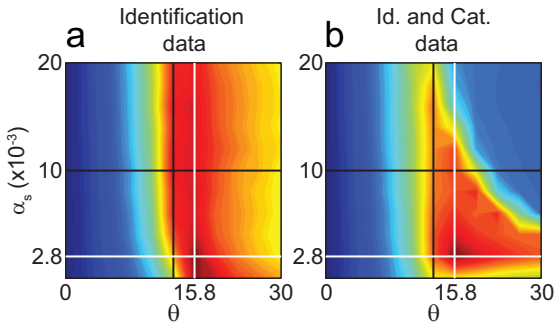

Figure 5.5. Fitting landscapes for learning parameter when considering likelihood functions from Equation 3.12 (a) and 5.16 (b). **(a,b)** Fitting landscapes for α versus DDM threshold θ_0 for identification data before (a) and after accounting categorization data (b). Color map indicates minimum log likelihood in blue and maximum in red. White lines indicate the parameters that best fit all data. Note that the landscape is not well constrained for identification data only (a), but when accounted the categorization data α has now a unique solution (b).

best estimates highlighted in white). Best fitting parameters can be found in Table 5.1.

5.4.3 Fitting results

As highlighted in the previous section, we found that it was now possible to fit simultaneously the psychometric and chronometric curves of both identification (Figure 5.7a,b) and categorization (Figure 5.7c,d) tasks with a single set of parameters (Table 5.1). Because the RL-DMM has one additional parameter, we conducted a Bayesian information criterion (BIC) analysis, which measures fitting quality with appropriate penalties for the number of parameters (Schwarz, 1978). The BIC model comparison showed that the RL-DDM with stimulus-dependent RL had superior performance to the vanilla DDM and its various variants (Figures 3.5 and

5.2) except the case of DDM with random weight noise (Figure 3.8; Table 5.1).

Model	Procedure	k	β	θ_0	τ (ms)	t_{ND} (ms)	l_r (%)	α	σ_w	BIC
DDM	**Fit Id.**, pred. Cat.	1.12×10^{0}	0.479	11.1	441	274	3.06	n/a	n/a	2208
DDM	**Fit Cat.**, pred. Id.	7.86×10^{-2}	0.580	10.0	649	220	2.37	n/a	n/a	3628
DDM	**Simul. fit**	2.18×10^{-4}	0.213	14.9	505	180	3.00	n/a	n/a	2848
DDM+bias	**Fit Id.**, predict Cat.	1.92×10^{0}	0.604	10.7	344	285	2.77	n/a	n/a	2763
RL-DDM	**Fit Id., α fit to Cat.**	1.92×10^{0}	0.604	10.7	344	285	2.77	9.81×10^{-3}	n/a	170
DDM+random	**Fit Id., σ_w fit to Cat.**	2.43×10^{0}	0.651	10.1	820	280	4.54	n/a	0.409	132

Table 5.1. Fitted parameters and comparative goodness-of-fit for all versions of DDM used in this thesis.

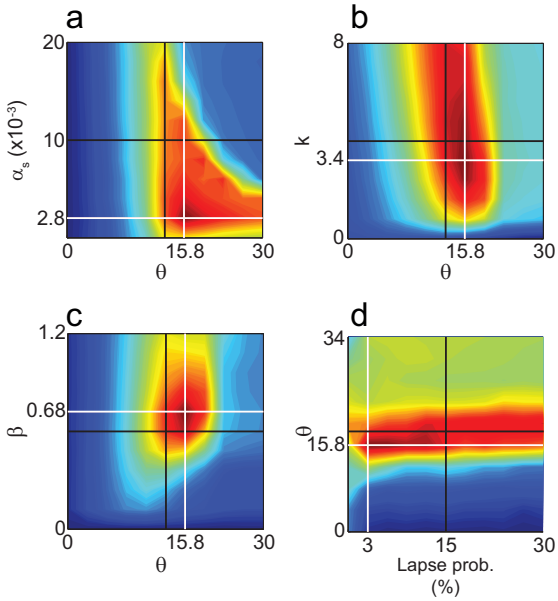

Figure 5.6. Fitting landscapes for RL-DDM that maximize likelihood function from Equation 5.16. Color map indicates minimum log likelihood in blue and maximum in red. White lines indicate the parameters that best fit the data (Table 5.1).

5.4.4 Model predictions

The RL-DDM also generated a prediction for the shape of history-dependent choice bias ("updating") for both tasks (Figure 5.7e,f). It is important to note that these functions were not fit, since the only data used for the fits were the trial-averaged accuracy and RT curves. Remarkably, the predictions of the model closely matched the data both qualitatively and quantitatively. For the categorization task, as expected, the model captured the strong dependence of the updating curve on stimulus difficulty (Figure 5.7f). This is due to the fact that for easy stimuli the predicted probability of a rewarded trial (i.e. the value of the evidence at stopping time) is nearly equal to 1 and there is little surprise and thus reduced learning. For the identification task the model was also able to capture the relative lack of stimulus dependence of the updating curve (Figure 5.7e). This is explained by the fact that there is a larger contribution of the stimulus-independent updating term (b) to choice bias, given the low stimulus concentrations in this task. This is in agreement with RL not affecting the model's performance in identification (Figure 5.4), suggesting that identification is less susceptible to this form of variability, i.e. category boundary fluctuations. An intuition for such an effect is provided further below. These effects were also observable when considering the full updating curves (Figure 5.8).

We have shown that ongoing changes of stimulus weights due to learning (equivalent to changing the categorization boundary) is capable of explaining both the accuracy and RT data on two tasks. But is learning necessary? Trial-by-trial random weight fluctuations can explain and actually outperform RL-DDM on the account of average behavioral data (Table 5.1). However, this version of the model has uncorrelated trial-by-trial noise, failing to predict the reward and stimulus dependent updating curves (Figure 5.10) and can therefore be rejected.

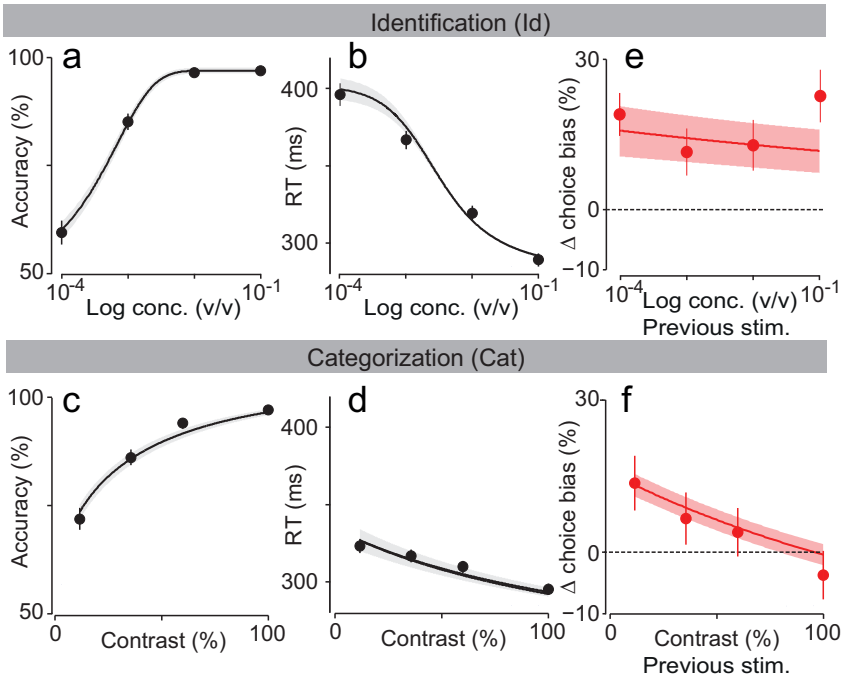

Figure 5.7. DDM with bias and stimulus learning explains identification and categorization task simultaneously. **(a,b,c,d)** Choice accuracy (fraction of correct trials) and odor sampling duration in identification task (a,b) and categorization task (c,d). Solid black line represents model with 6 parameters from the DDM fitted to identification task, plus the learning term, α_s, that was allowed to change in order to fit the categorization task. **(f,g)** Changes in choice bias conditional on rewarded previous trial stimulus, for identification (f) and categorization (g) tasks. Red points correspond to behavioral data, and solid red line to the predicted change from the model fitted to (a-d). Shaded red area corresponds to the 95% confidence interval of the model when considering the observed number of trials in the behavioral data.

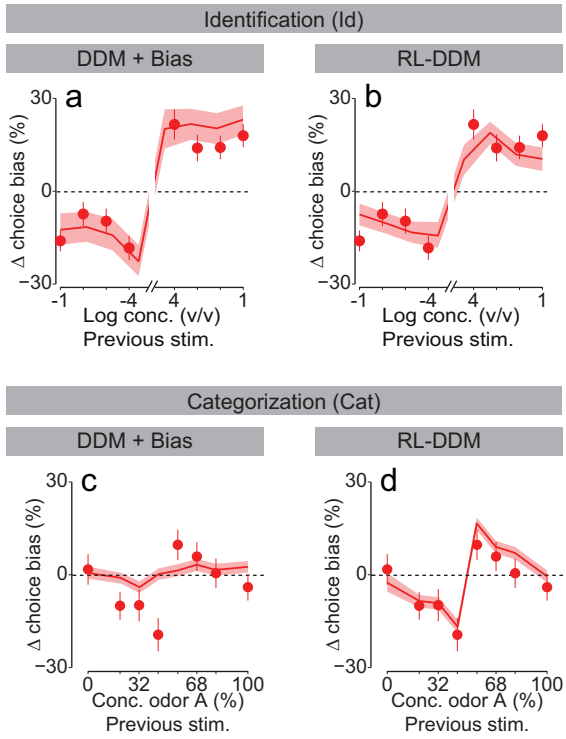

Figure 5.8. DDM+bias and RL-DDM predictions for full updating curves. **(a,b,c,d)** Change in choice bias after a correct choice and a given stimulus. Data presented for identification (a,b) and categorization tasks (c,d). Solid lines represent the predictions for DDM with bias (a,c) and RL-DDM (b,d). Error bars show SEM across rats' means. Shaded areas the predicted SEM from the model considering the same number of trials as observed in the data.

The RL-DDM model not only predicts changes in bias on a trial-by-trial basis but also changes in RTs. By changing the integration weights we effectively changed the drift-rate and starting position of integration. A rewarded difficult stimulus should originate an increase of μ in the rewarded direction. This implies an increase of choice bias accompanied by a decrease in integration time necessary to hit the same bound.

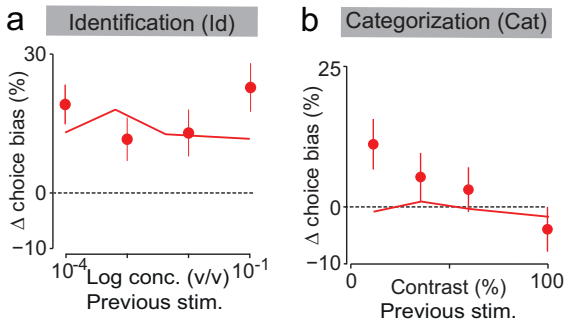

Figure 5.9. DDM+bias with random weights fluctuations predictions for changes in choice bias. **(a,b)** Changes in choice bias after a correct choice and a given stimulus for both identification (a) and categorization (b) tasks. Solid lines represent the predictions for DDM+bias with random weights (Figure 3.8). Error bars show SEM across rats.

We tested this prediction by looking at the change in RTs following a reward and a given stimulus difficulty in a similar way to what was done for choice bias (Figure 5.9). The obtained curves did show qualitative similar results despite not as striking as what was seen for choice bias. The model predicted that difficult rewarded stimuli should be followed by faster decisions towards the same direction in both tasks.

As a final test of the RL-DDM we assessed whether it could predict the behavioral results for all the intermediate concentration levels of the mixture task. To do so, we fitted the model to the 8 pure odor stimuli (equivalent to the identification task) and the additional 6 highest concentration mixtures (equivalent to the categorization task). We then tested whether this model could predict the remaining 18 points (Figure 2.5a). Thus, we required the model to generalize to a novel set of stimuli on which it was not trained, a rigorous form of cross-validation.

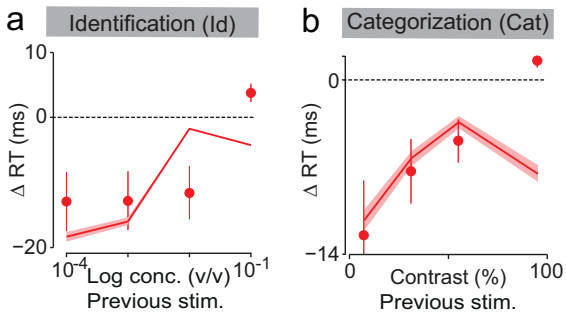

Figure 5.10. RL-DDM predictions for changes in reaction times. **(a,b)** Changes in RTs after a correct choice and a given stimulus for both identification (a) and categorization (b) tasks. Solid lines represent the predictions for RL-DDM. Error bars show SEM across rats. Shaded areas the predicted SEM from the model considering the same number of trials as observed in the data.

We found that the model was indeed able to predict well the full set of psychometric (Figure 2.5b, solid lines) and chronometric (Figure 2.5c, solid lines) functions.

These results show that the rats use the same decision-making network to solve the two tasks. The introduction of the learning rule allowed explaining what happens to categorizing mixtures of decreasing concentrations. The crosstalk between stimulus noise and boundary uncertainty is thus essential to understand the interplay between accuracy and reaction times in these tasks.

5.4.5 Analysis of category bound fluctuations

Finally, we sought to use the model to gain insight into how RL works in conjunction with integration-to-bound to explain the difference between identification and categorization task performance. To do so, we analyzed

the dynamics of the weights in relationship to sensory evidence (Figure 5.11).

For each simulated trial, we inferred the mean drift rate by considering the integrated evidence and integration time. As any model reaches a decision, it has access to two variables: amount of evidence at the bound and the decision time t_d. For better understanding the dynamics immediately before the multiplication of the weights, we looked at the combination of sensory evidence (s_1, s_2) for each simulated trial. For trial j there is a noisy sensory evidence trajectory (integration layer). This means that by the end of trial j we can compute the mean drift rates that gave rise to a particular decision:

$$\left\langle \mu_i^j \right\rangle = \frac{s_i^j}{t_d^j - t_r} \tag{5.17}$$

First, we plot simulations for the DDM model that was fit to the identification task and outperformed in categorizing stimuli (Figure 5.11a,b; solid black lines from Figure 3.5). Here, each dot is a simulated trial and the scatter of dots around each stimulus reflects the impact of stochastic noise in the DDM. Each group in Figure 5.11a,b has been segregated taking into account the Mahalanobis distance (Mahalanobis, 1936), as each line represents the distance of $D = 1$ for a particular stimulus set.

The category boundary in this ideal DDM is a line with slope equal to 1 with all stimuli below this line "left" choices and all stimuli above this line as "right" choices. In Figure 5.11c,d we plot for this model the most difficult "right" decision stimuli for each task (correct choices in blue and errors in red). In this model with ideal category boundary, it can be seen that performance on the categorization task is expected to be much higher than for the identification task.

We now considered the integrated evidence of Equation 5.10 and combine it with the choice function of Equation 3.8. We see that:

$$w_1 s_1(t) + w_2 s_2(t) + w_b = 0 \tag{5.18}$$

should represent the separation line between the two stimuli, and thus we can rewrite Equation 5.18 as:

$$s_2(t_d) = -\frac{w_1}{w_2} s_1(t_d) + \frac{w_b}{w_2} \tag{5.19}$$

Considering the straight-line equation $y = mx + i$ we see that in our integrated evidence plots the boundary separation can be drawn with slope $m = -w_1/w_2$ and intercept $i = w_b/w_2$.

Stimulus weight fluctuation should then have an impact in the slope of the boundary line separating the classification between left and right stimuli, and w_b should influence the origin intercept on that stimulus representation (Figure 5.11).

In Figure 5.11e,f we show the effect of fluctuating weights (variability in the category boundary) in the RL-DDM (grey area indicates 1 standard deviation from the mean). It can be seen that for the identification task boundary fluctuations changed the classification of very few trials, but in the categorization task many were affected (grayed points unchanged, red/blue points changed).

The difference in effects can be understood by considering that in the RL-DDM stimulus weights have a multiplicative effect on evidence strength. Thus, weight fluctuations correspond to rotations around the origin; the effects are larger for larger stimulus values. Therefore, the difficult mixture stimuli of the categorization task, which have higher concentrations, are much more susceptible to these fluctuations. It can also be seen that the amplitude of weight fluctuations is larger for the categorization task, but this effect cannot be primarily responsible, as it is

not present in the mixture identification task (in which all stimuli are interleaved).

As we can see, stimulus fluctuations have a larger impact in mixture categorization than in odor identification. However, we wanted to test if our weight changes are reflecting rats' biases at the local level. To do so we calculated "local bias" as the bias obtained in a 10 trial sliding window in order to quantify how these biases change over the course of one session (Figure 5.12). We did the same for RL-DDM weights by calculating the difference in $\Delta w = w_1 - w_2$. By letting the RL-DDM go through the same trial history as experienced by the rats, we estimated how far from the ideal bound should the model be. We plotted an example of how local bias and weights fluctuated over the course of a particular session (Figure 5.12a). We now tested how well our model predicts the changes in bias by calculating the cross-correlation between the two curves (Figure 5.12b). We saw that weights were a good predictor of local bias, as the lag in the number of trials peeked at 1 with a slow decay over time.

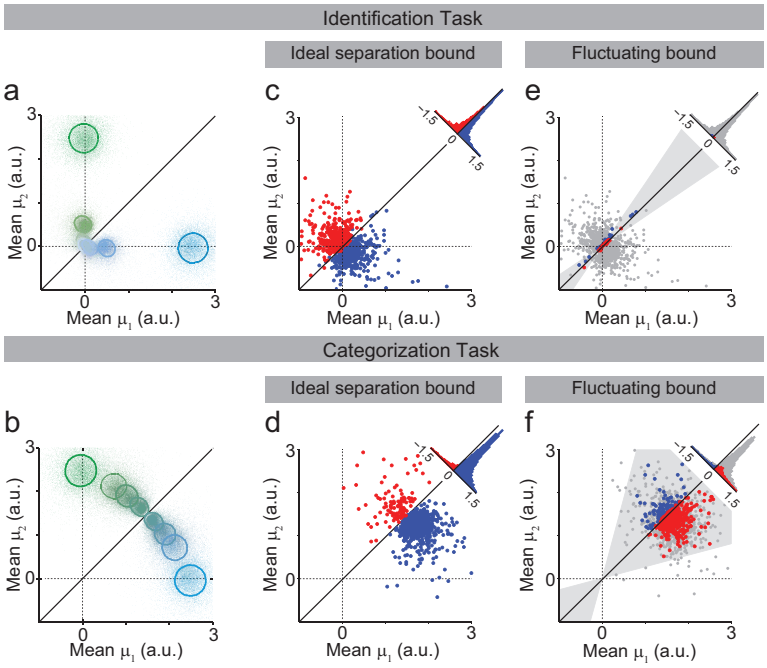

Figure 5.11. Weight fluctuations amplify errors in categorization task. **(a,b)** Stimulus space for identification and categorization task. Each point represents a combination of mean drift rates for a given trial in the pure DDM that outperformed in the categorization task (Figure 3.5). Solid oval lines represent the Mahalanobis distance of 1 in relation to the average for each of the eight stimuli. Unity line depicts the ideal classifying process. Plotted are trials for both identification (a) and categorization (b) task. Color code follows the same logic as Figure 2.2. **(c,d)** Mean drift rates for the most difficult left-choice in the case of the RL-DDM model. Blue signals the correct classified choices and red the incorrect. Histograms quantify the difference $\langle \mu_1 \rangle - \langle \mu_2 \rangle$. **(e,f)** Same as (c,d) but now with fluctuating weights depicted as the slope of the category bound $s = -w_1/w_2$. Grey area indicates 1 standard deviation of weight fluctuations for both identification (e) and categorization (f) tasks. Blue indicates trials that were originally incorrect in (c,d) but became correct, and red indicates incorrect trials that were originally correct. Light grey dots indicate answers that remained unchanged. Histograms quantify the population for each of the three groups.

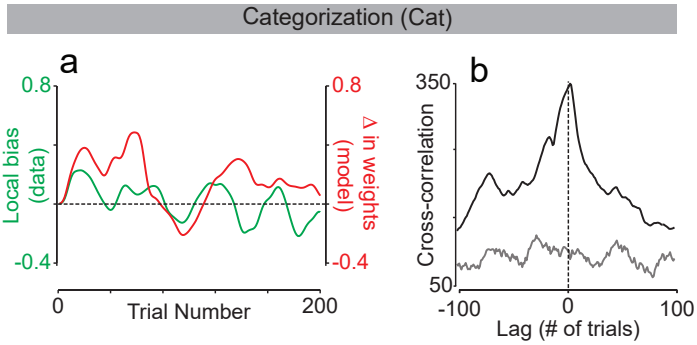

Figure 5.12. Model weights correlate with rats' local bias. **(a)** Local bias (green) and model's difference in weights (red) as a function of trial number within an example session. We calculated these variables by considering a 10 trials sliding window. **(b)** Cross-correlogram of local bias vs Δw. Correlation peeks at 1 trial lag and slowly decays over trials. Solid grey represents the highest correlation expected cross-correlogram with random shuffles of Δw.

5.5 The two-race accumulator model

We also tested alternative versions of our RL-DDM in which we controlled for the accumulator process that preceded the decision layer (Smith & Vickers, 1988; Usher & McClelland, 2001; Kepecs, Uchida, & Mainen, 2008). Alternative models based on confidence as a difference between two alternatives could also reproduce the relationship between difficulty and evidence at stopping time, but were not all explored here as they all pointed towards the same conclusion as RL-DDM. Nevertheless, we focus here in one particular version of the model which might shed light on the neural architecture behind confidence modulated updating (Lak et al., 2014; Costa, 2015; Kepecs, Uchida, Zariwala, & Mainen, 2008).

The two-race accumulator model (2RM) is in all ways analogous to the DDM except for the fact that evidences for both "left" and "right" are not

combined before the decision layer (Figure 5.12a). This model quantifies the absolute evidence contrasting with the relative computation done in DDM (Ratcliff & McKoon, 2008; Usher & McClelland, 2001). Thus, for this version we now have two integration particles:

$$e_i = \vec{w}_i \int r_i(t)dt \pm w_b \tag{5.20}$$

The two races now compete and the decision made will correspond to the race that first reaches criterion θ. It is important to note that this model can be made equivalent to our RL-DDM due to the incorporation of a collapsing bound (Churchland et al., 2008; Gold & Shadlen, 2007). A collapsing bound can be thought of as an urgency signal that forces the DDM to commit to a faster decision by sacrificing the magnitude of the evidence difference.

For 2RM we implemented a different learning rule that maintained the same flavor as our previously implemented delta rule.

$$\begin{aligned} \Delta\vec{w}_i &= \alpha\big(\lambda - f(\vec{w}_i\vec{r})\big) f'(\vec{w}_i\vec{r})\vec{r} \\ &= \alpha(\lambda - y)y(1-y)\vec{r} \end{aligned} \tag{5.21}$$

in which y is the logistic function that we used to apply the gradient descent learning while fitting our data iteratively. It can be though of as the predicted outcome expected by our model and depends on the difference between evidences Δe (highlighted by the red line in Figure 5.12a).

$$y = f(\Delta e) = \frac{1}{1 + \exp\big(\Delta e\big)} \tag{5.22}$$

We followed a similar fitting procedure as what was done for the DDM: first we fitted the integration and stimulus dependent parameter do the identification task and then allowed the learning rate to change in order to

fit the categorization data. As we can see from Figure 5.12b-e we obtained similar results to what was seen for RL-DDM.

This model demonstrates that the integration process can be implemented in a variety of ways and reemphasized that confidence based reinforcement learning is of paramount importance to explain mixture categorization performance.

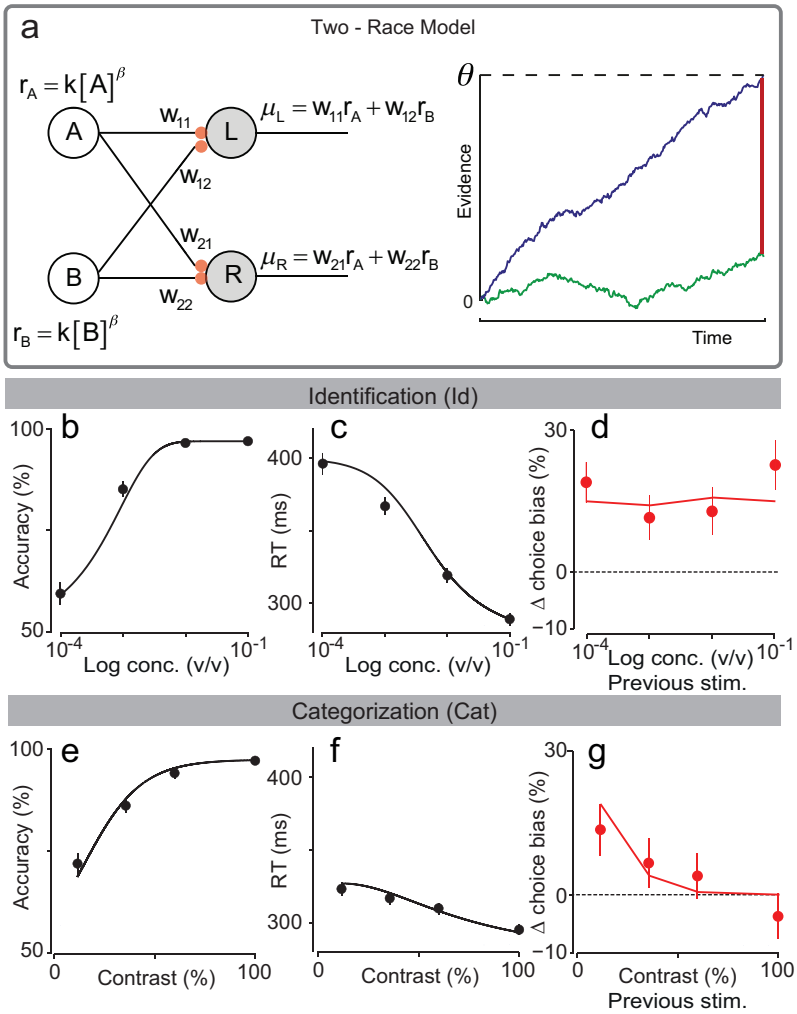

Figure 5.13. Two-race model with learning also explains identification and categorization task simultaneously. **(a)** RL-2RM is analogous to RL-DDM with the only difference existing at the decision particle – in this case two simultaneous races occur and the decision is made when one hits the threshold. Weights are updated every trial depending on the difference between evidences (highlighted in red) and according to Equation 5.21. **(b,c,d,e,f,g)** Model's fits (b,c,e,f) and predictions (d,g) for both identification and categorization task.

Chapter 6

Neural signatures of weight updating

"Accept change"
– Track changes option,
Microsoft Word

6.1 Chapter Summary

In the previous Chapter we saw that RL was essential to explain our behavioral data. In particular, the inclusion of this rule in a drift-to-bound model allowed us to predict and explain the changes in choice biases observed in the data. We further validated our model by showing that it was able to predict a completely new set of data, the *hybrid* odor mixture identification task.

In this Chapter, we look into neural signatures of RL-like updating. For that we study what type of changes we expect to see in neural data if our model is correct. We tested modulation of previous outcome and difficulty in firing rates by taking Generalized linear models and fitting them to our neural data. This chapter is divided in three sections:

- **Introduction** – a detailed introduction about the neural prediction from RL-DDM is given. Additionally, we introduce our candidate areas for such representations - the PIR, OFC and VS. Lastly, we give a brief background on the neural data used in this Chapter and how it was collected.

- **The effect of trial outcome in OFC and VS** – in this section we implement a GLM that estimates the relative influence of trial outcome in OFC and VS while rats perform mixture categorization. We demonstrate that rewards modulate the activity of these neurons.

- **Stimulus updating in OFC** – where we further explore the interaction of stimulus and rewards in our GLM fits. We see that OFC is modulated by previous trial outcome and difficulty.

6.2 Introduction

Rats are exposed to two different sources of uncertainty while discriminating odors (Chapters 2 to 5). We showed that one of these sources is related to uncertainty in regards to the category bound separation, the iso-concentration line of 50% contrast. We were able to grasp this idea by introducing a deterministic rule of RL, showing that these results can be explained by not adding a new source of noise but an apparent sub-optimal strategy (Beck et al., 2012).

We now ask the question of how this might be implemented in the brain. The DDM is a good outline to understand how noisy inputs can tamper with decisions. It also permits quantitative predictions that are easily tested and have successfully been used to explain a variety of complex behaviors (Milosavljevic et al., 2010; Ratcliff & McKoon, 2008; Ratcliff et al., 1999; Krajbich et al., 2012; Mulder et al., 2012; Krajbich & Rangel, 2011; Ditterich, 2006; Ratcliff & Smith, 2004; DasGupta et al., 2014; Ratcliff & Rouder, 1998; Bogacz et al., 2006; Bowman et al., 2012; Gold et al., 2008; Usher & McClelland, 2001; Drugowitsch et al., 2012; Lak et al., 2014; Hanks et al., 2015; Kiani et al., 2008; Churchland et al., 2011; Krajbich et al., 2010; Beck et al., 2008; Laska et al., 2004; Palmer et al., 2005; Gold & Shadlen, 2007; Ratcliff, 1978; Smith & Vickers, 1988; Hanks et al., 2014; Churchland et al., 2008; Brunton et al., 2013). However, a diffusion particle treats excitation and inhibition symmetrically, has unlimited negative values and originates responses dependent of two thresholds. All these facts challenge the concept of a direct implementation by a neural network (Usher & McClelland, 2001; Ratcliff & Smith, 2004; Gold & Shadlen, 2007). It also brings forward the issue of how to implement decisions with more than two choices (Churchland et al., 2008). Nevertheless, DDM can be thought of as a theoretical abstraction that allows testing of clear predictions (Palmer et al., 2005). Additionally,

the DDM can be set up in a variety of different hypothesized neural architectural implementations (Gold & Shadlen, 2007; Hanks et al., 2015). In fact, throughout this thesis, we tried to maintain a layered structure for our computational models so that we could shed light on how these might be implemented in the brain and respective neural network. The more closely related and neuron inspired 2RM (Usher & McClelland, 2001) showed us that the important question is not so much how DDM might be working in the brain, but related to identifying which areas might be contributing to the apparent implementation of sensory, integration and decision layers.

6.2.1 Neural predictions of RL-DDM

Our RL-DDM model has clear predictions of what should be occurring at the integration process and the combined evidence particle of Equation 5.9. The average path for a given stimulus has a particular drift rate μ. This average path is highlighted by the black arrow in illustrative Figure 6.1a. Let us consider a trial in which a given drift rate originated a "left" decision. If this decision is correct, our model predicts an increase in weights from feedback of Equation 5.12, which then reflects an increase in drift rate for the following trial, coherent with the idea of more frequent and faster left choices (Figures 5.8 and 5.10). Conversely, an error should have the opposite effect (Figure 6.1). What are the implications and predictions for our neural data?

The first prediction is that the untampered sensory representations (before weight multiplication) should be unaltered by this outcome dependency. The second prediction is that downstream, when sensory representations are transformed into action values, these evidence accumulation neurons should present a neural activity change in the direction of the rewarded side. If a unit has a given drift rate for a particular bound then it should increase its neural activity after a reward. Conversely it should

decrease as it receives an omission (Figure 6.1b, blue path represents a correct trial, red path an incorrect one).

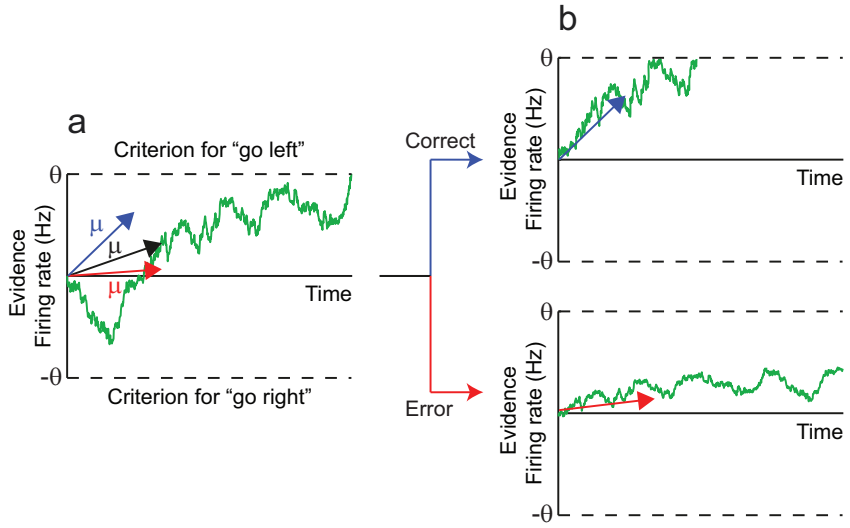

Figure 6.1. RL-DDM predictions for change in evidence rate after trial outcome. **(a)** A given stimulus has an average drift path that defines drift-rate μ, i.e. the mean rate of evidence for that particular stimulus and choice. **(b)** RL-DDM predictions after a given decision: if a decision is followed by a reward one expects to see an increase in μ and thus an increase of evidence (firing rate). Conversely the opposite effect should occur for an error trial. Blue highlights the change after a correct trial, red an incorrect trial, black the mean response and green an individual trial race contaminated by noise.

Unfortunately, we were not able to collect electrophysiological responses from our subjects in the scope of this study (Vicente, 2015). Nevertheless, previously collected data for the categorization task might allow us to test these predictions (Miura, Mainen, & Uchida, 2012; Costa, 2015).

Piriform cortex (PIR) is the main olfactory processing cortex (Haberly, 2001; Rennaker, Chen, Ruyle, Sloan, & Wilson, 2007; Wilson, 2003; John-

son, Illig, Behan, & Haberly, 2000; Wilson, 2000). It has been shown to be a highly associational cortex that identifies odorant mixtures in an apparent sparse representation (Stettler & Axel, 2009). However, one can easily use PIR electrophysiological information to decode odor identity (Miura et al., 2012). In fact, Miura et al. recorded from areas of anterior PIR (less associational than posterior PIR) while rats performed the mixture categorization task. They observed that taking into account a reduced number of cells (approximately 50) one could easily outperform the behavioral accuracy of the rats. This is in agreement with what we would predict from our RL-DDM model. But lack of observation does not mean proof of concept (Chalmers, 1999). Ideally, we aimed to test if changes of evidence accumulation occur after trial outcome feedback.

6.2.2 Time wagering task, the OFC and VS

The orbitofrontal cortex (OFC) and ventral striatum (VS) are two brain regions implicated in behavioral supervision and outcome evaluation (Wallis, 2007; Feierstein et al., 2006; Wallis & Miller, 2003; Padoa-Schioppa & Assad, 2006; Hikosaka & Watanabe, 2000; Roesch, Stalnaker, & Schoenbaum, 2006; Murray, O'Doherty, & Schoenbaum, 2007; Bowman et al., 2012; Illig, 2005; Lak et al., 2014; Mulder et al., 2012; Simmons et al., 2007; Grinband, Hirsch, & Ferrera, 2006; Stages, 2002; Lak et al., 2014; Changeux & Dehaene, 2008; Lau & Glimcher, 2005; Green, Benson, Kersten, & Schrater, 2010). In particular, for the categorization task, neurons in the OFC were found to encode information about presented stimuli, in the period of time before a decision was made (Feierstein et al., 2006). In the same study, OFC neurons were also encoding choice direction, when the animal was moving towards the choice port. These findings are in concordance with the view of OFC as playing a central role in goal monitoring, shown in the context of reward-based decision making (Balleine & Dickinson, 1998; Burke, Franz, Miller, & Schoenbaum, 2008;

Fellows, 2011; Morrison & Salzman, 2011; Padoa-Schioppa & Assad, 2008; Padoa-Schioppa & Cai, 2011; Roesch, Taylor, & Schoenbaum, 2006; Schoenbaum et al., 2011; Schultz, Tremblay, & Hollerman, 2000; Takahashi et al., 2013; Wallis, 2012). OFC activity was also found to be encoding decision confidence in rats performing the odor categorization task (Kepecs, Uchida, Zariwala, & Mainen, 2008).

Sharing a similarly important role in evaluation of performance is the VS (Botvinick et al., 2009). This area was found to interact with OFC to guide optimal courses of actions that ultimately lead to rewards (Hare et al., 2008; McDannald et al., 2011; Simmons et al., 2007). Moreover, it is very likely that activity in VS neurons also correlates with decision confidence (Daniel & Pollmann, 2012; Hebart, Schriever, Donner, & Haynes, 2014).

Costa and Mainen developed a similar behavioral setup that allowed testing these issues of reward expectation, magnitude, confidence and uncertainty in both behavior and neural responses (Costa, 2015). In particular, they developed a waiting time wagering task (Figure 6.2a) that was in all ways similar to our categorization task except for the fact that rewards were largely delayed after the choice was made. This randomized delay, combined with catch trials, was used to test how confident rats would be after committing to a particular response (Kepecs, Uchida, Zariwala, & Mainen, 2008; Lak et al., 2014; Costa, 2015). They found that waiting time was correlated with stimulus difficulty, predicted outcome and choice, indicating that rats were wagering the level of certainty of their own perceptual decisions. Additionally, they found that these behavioral variables were correlated with neural activity in OFC and VS, suggesting that these areas code the value of each port adaptively.

Costa and Mainen also introduced the concept of blocked trials in mixture categorization (Figure 6.2b). This was done in order to manipulate reward size and identify its effects in confidence reports. They showed

that reward magnitude manipulation changed rats' biases towards the most rewarded side (Figure 6.2c), although this did not change the rats' confidence report (Costa, 2015). This change in choice bias with reward size can actually be predicted by the RL-2RM model (Figure 6.2), by changing the slope and bias of the logistic function of Equation 5.22. The solid lines in Figure 6.2c depict fits obtained for this particular version of the task. Nevertheless, in this study we will focus on the effect of reward and not reward size. More information regarding reward size can be found in Costa's PhD thesis (Costa, 2015).

Neural activity from OFC and VS cells was recorded in 6 rats performing this version of mixture categorization (Figure 6.3). Using chronically implanted tetrodes, 59 and 64 units were recorded from OFC and VS, respectively. Neurons showed varied responses to different events of the task, but a significant group in both areas showed outcome relative preference (VS: 25% of cells, $p < 0.05$; OFC 22%, $p < 0.05$, permutation tests).

For the purposes of our original RL-DDM expectations in neural activity (Figure 6.1b), we hypothesized that this dataset might show neural signatures of outcome weight modulation, as RL implies an adaptive change in w_i and w_b which in turn implies a change in relative value between the two choice ports. To do so, we decided to take a generalized linear model (GLM) fitting approach and ask how well different task parameters can fit fluctuations of neural activity.

6.3 The effect of trial outcome in OFC and VS

We aimed to test whether neural activity in both OFC and VS reflected weight updating on a trial-by-trial basis. To do so we considered all recorded units from 6 rats while performing the categorization task and divided the analysis in two groups, one for OFC and the other for VS. To see effects of trial-by-trial influence we developed GLMs. In statistics, the

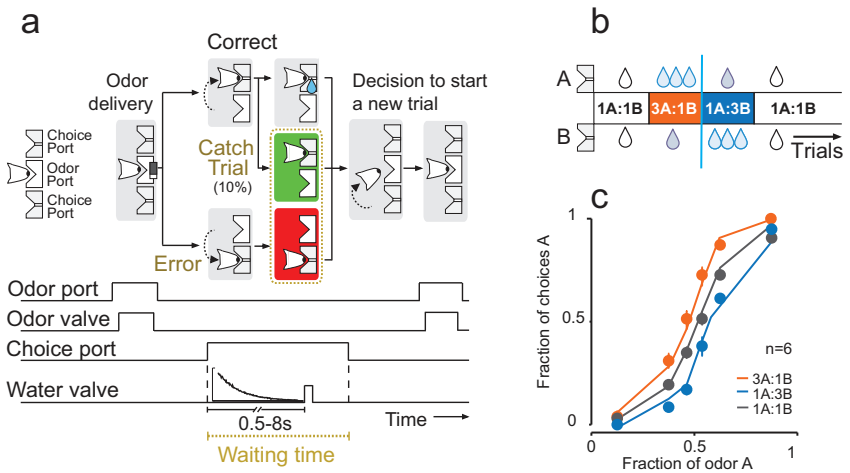

Figure 6.2. Time wagering confidence task. **(a)** Schematic representation of the behavioral paradigm. All task variables were comparable to the categorization task except for the delay in water delivery. This pseudo-random delay was drawn from an exponential function with decay of 1.5 s, a 0.5 s offset and a maximum of 8 s. Catch trials consisted of 10% of correct trials. The amount of time rats were willing to wait was considered the proxy for confidence report. **(b)** Reward magnitude manipulation. Each session started with an unbiased block and was followed by two other blocks in which one side was rewarded 3 times more. Each block consisted of approximately 130 trials. **(c)** Effects of reward magnitude on performance. Psychometric curves divided by block. Solid lines are fits obtained from a RL-2RM model (Figure 5.13). Error bars are SEM over 6 rats. Adapted from Costa (2015).

GLM is a flexible generalization of ordinary linear regressions that allows for response variables that have error distribution models other than a normal distribution. GLMs have been used to describe spike trains in a variety of situations (Gerwinn, Macke, & Bethge, 2010; Pillow et al., 2008; Truccolo, Eden, Fellows, Donoghue, & Brown, 2005; Chichilnisky, 2001). A notorious usage of such models is the linear-nonlinear Poisson cascade model (Simoncelli, Paninski, Pillow, & Schwartz, 2004; Chichilnisky, 2001).

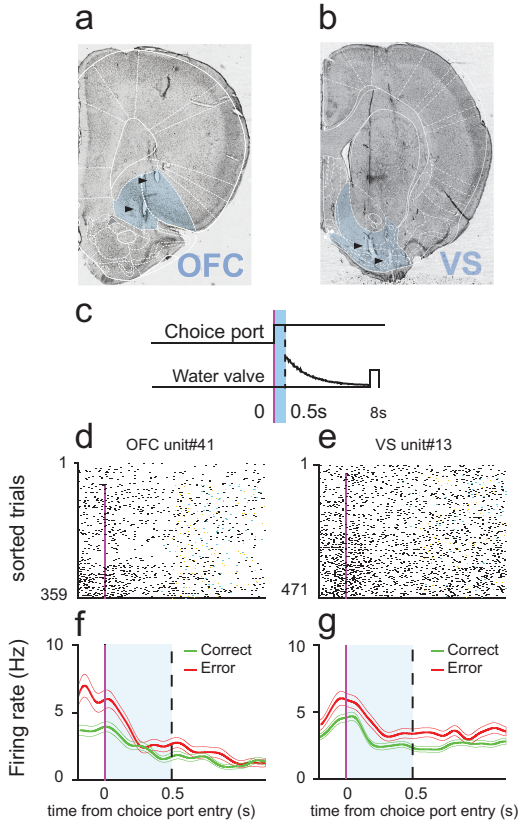

Figure 6.3. OFC and VS recordings. **(a,b)** Example Nissl-stained coronal sections showing electrolytic lesion sites (black arrows) from tetrodes located in OFC (a) and VS (b). **(c)** Timing of outcome anticipation period. Neuronal activity was aligned to each trial choice port entry, signalled by the first break of photo-beams within each choice port, and analyzed for 500 ms after choice port in - light blue bar. **(d,e)** Raster plots represent neural activity, with each row corresponding to a single trial and each tick mark to a spike. Magenta marks signal the time of choice port entry and rows with no mark are initiated trials where no choice was made. **(f,g)** Peri-event time histogram (PETH) of example neuronal units. Trials are grouped by outcome – correct (green) and error (red). Lines represent mean ± SEM over trials. Adapted from Costa (2015).

For this particular case, we opted for a very simplistic approach. We considered a simple GLM with noisy Poisson generators as the link function and calculated estimators that might explain changes in firing rates.

6.3.1 Generalized linear model implementation

Our first point of interest was to check for the effect of previous trial reward in current trial neuronal activity. To do so we considered the average firing of each unit from OFC and VS. We calculated the total number of spikes, s, in one trial, t, divided by the time length of that same trial, ΔT (time from odor poke entrance to choice poke out, Figure 6.2a).

$$f(t) = \frac{s_t}{\Delta T_t} \tag{6.1}$$

We defined choice variable c. This variable signals the rat's choice as a 1 (left) or -1 (right). We also defined reward outcome variable λ. This variable consists of the interaction between choice, $c(t)$, and reward, $r(t)$. So, for a given trial, reward outcome is:

$$\lambda(t) = c(t) \times r(t) \tag{6.2}$$

λ will signal the presence of a reward on the left if equal to 1, -1 presence of reward on the right and no reward for 0.

We aimed to see if firing rates were modulated by choice (and thus reflecting poke value) and previous reward outcome. To do so, we fitted the following GLM with Poisson noise for each neuron:

$$f(t) = \beta_0 + \beta_1 c(t) + \beta_2 \lambda(t-1) \tag{6.3}$$

in which β_1 is defined as the choice side parameter and β_2 the previous outcome parameter.

From our RL-DDM predictions we expected to see a strong correlation between β_1 and β_2. The rationale for this is the following: if RL is occurring adaptively (updating) as stated before, then a unit that is coding the value for a particular choice should reflect that adaptive change in value in a trial-by-trial manner. Thus, if a unit presents a strong side preference throughout the course of one trial (and thus $\beta_1 \neq 0$) then it should present an equivalently strong dependency for previous trial outcome ($\beta_2 \neq 0$) and in the same direction as β_1.

6.3.2 GLM results

We fitted 59 OFC and 64 VS units and obtained the equivalent number of estimators (Figure 6.4a,b). We saw that for both areas choice side preference and previous outcome parameters were correlated (for OFC, Spearman's $\rho = 0.311$, $p < 0.05$; and for VS, Spearman's $\rho = 0.479$, $p < 0.01$). These results demonstrated that neurons in OFC and VS that are correlated with choice bias are also modulated by the outcome of the previous trial.

However, we saw that a reduced population showed significant values for both parameters. For OFC, 7% of the cells showed significant modulation of choice (4/59, $p < 0.05$ for β_1; Figure 6.4c) and 14% showed changes in firing rate for previous outcome (8/59, $p < 0.05$ for β_1, Figure 6.4d). Of these, only 2 units showed significance for both estimators (black points in Figure 6.4a). For VS, 8% of the cells showed significant modulation of choice (5/64, $p < 0.05$ for β_1; Figure 6.4e) and 16% showed changes in firing rate for previous outcome (10/64, $p < 0.05$ for β_1, Figure 6.4f). In VS, the number of overlapping units were 2 (black points in Figure 6.4d). Possible reasons for such reduced numbers are discussed in the last Chapter of this thesis.

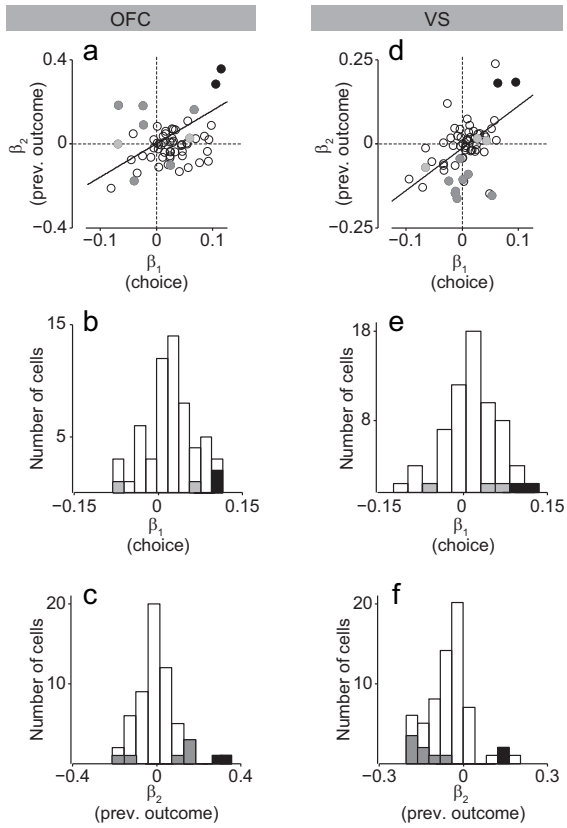

Figure 6.4. Average firing rate GLM parameters for current choice and previous trial outcome. **(a)** Previous outcome estimator, β_2, as a function of choice side estimator, β_1, for OFC neural data. **(b,c)** Histograms for estimator parameters β_1 (a) and β_2 (b) and respective number of cells for each bin. **(d,e,f)** Same as (a,b,c) but for VS data. Black indicate cells that show significance ($p < 0.05$) for both estimators, light gray for β_1 and dark grey for β_2. Dark lines in (a,d) represent best linear regressions for the plotted datasets.

6.3.3 Example cells

We now examined units that showed a large effect of both β_1 and β_2. For that we looked at the strongest effect (taking into account the sum of both p-values) for both OFC (cell #39) and VS (cell #23). Visual examination of the obtained cells show that they have slightly different profiles. The full peri-event time histograms (PETH) with 25ms time window profiles can be found in Appendix 2 and 3. We synthesize the observed patterns and profiles in the following paragraphs.

OFC cell #39 demonstrated integration from odor valve on until entrance in choice poke. When we conditioned the PETHs to trials that followed a particular outcome (reward vs no reward), we saw that neural activity went up following a rewarded trial when compared to mean activity (Figure 6.5, blue lines compared to black lines). Conversely, the same was observed for trials that were unrewarded for that side (Figure 6.5, red lines compared to black lines). The effect was sustained from start of stimulus presentation until choice port poke in.

VS cell #23 showed slightly different properties from the OFC cell. This cell fires throughout the course of the entire reward poke in until the rat pokes out. This cell might be reflecting the value of the choice port after a particular decision has been made, and thus used to evaluate the quality of that decision (Costa, 2015). We see that previous outcome changes the way this cell fires throughout the course of reward port events in the same direction: rewarded trials increased the firing rate for this unit and unrewarded trials decreased it (Figure 6.6).

These results apparently confirm our prediction of outcome modulation of evidence, value and thus neuronal activity. However, these modulations might be due to mere presence of rewards, i.e. a general bias that is reflected in the neural population. This is specially an issue if one considers that Costa and Mainen used blocked trials for reward. Thus, our observed changes in neural activities might be due to spurious influences of those

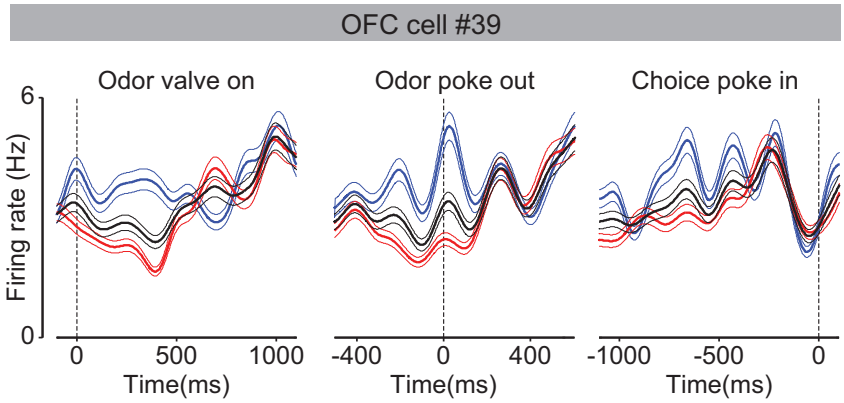

Figure 6.5. PETHs for example OFC cell modulated by previous outcome. Three plots are presented with alignment to three different events: odor valve on (beginning of stimulus presentation), odor poke out and choice poke in. Blue lines represent PETH done in trials that followed a rewarded left side choice. Red lines represent trials that followed an unrewarded trial or a right side reward. Black lines depict the average PETH. All lines are mean ± SEM over trials.

same trials. Still, our RL-DDM can make an even stronger prediction of how updating should influence neural activity in a trial-by-trial basis.

6.4 Stimulus updating in OFC and VS

As seen from Chapter 5, RL-DDM was implemented taking into account stimulus strength. We defined category uncertainty as the challenge to keep track of the true category boundary. As stimuli become closer to that categorical boundary, uncertainty regarding its true mapping should give rise to a larger effect of updating (Figure 5.7f). That is, our stimulus learning rule from Equation 5.12 dictates that learning should be larger when a stimulus is difficult than when compared to an easy stimulus.

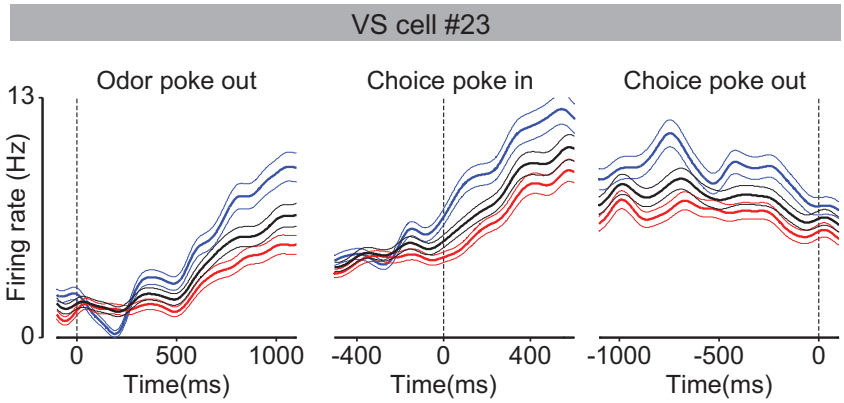

Figure 6.6. PETHs for example VS cell modulated by previous outcome. Three plots are presented with alignment to three different events: odor poke out, odor poke in and choice poke out (end of trial). Blue lines represent PETH done in trials that followed a rewarded left side choice. Red lines represent trials that followed an unrewarded trial or a right side reward. Black lines depict the average PETH. All lines are mean ± SEM over trials.

We tested this prediction by further expanding our GLM in order to incorporate stimulus difficulty and thus test if units in OFC and VS show outcome modulation of neural activity that is stimulus dependent.

6.4.1 Generalized linear model implementation

To further expand the GLM from Equation 6.3, we defined rewarded outcome variables that depend of stimulus difficulty. Thus, binary variable η indicates trials that were difficult (6% contrast), ν medium difficulty trials

(12%) and ε easy trials (60%). In this way we introduced more estimators to fit:

$$\begin{aligned} f(t) = \beta_0 + \beta_1 c(t) + \beta_2 \lambda(t-1)\eta(t-1) \\ + \beta_3 \lambda(t-1)\nu(t-1) \\ + \beta_4 \lambda(t-1)\varepsilon(t-1) \end{aligned} \tag{6.4}$$

By separating the trials into η, ν and ε, we estimated the relative influence of each type of trials with β_2, β_3 and β_4.

6.4.2 GLM results

We fitted 59 OFC and 64 VS units and obtained the equivalent number of estimators for difficulty and outcome modulation (Figure 6.7).

For OFC, correlation between the different estimators can be found in Table 6.1. The influence of previous stimuli in current firing rates was correlated with difficult stimuli. Conversely, medium and easy stimuli showed no correlation.

Estimators	ρ	p
β_1,β_2	0.350	**0.01**
β_1,β_3	0.193	0.14
β_1,β_4	0.165	0.21

Table 6.1. Spearman's rank correlation between GLM estimators for OFC data.

For VS, correlation between the different estimators can be found in Table 6.2. In this case, the effect of stimulus difficulty had no additional modulation in firing rates.

To better understand these effects, we compared the slopes of linear regressions done to the plots of Figure 6.7. The obtained slopes for OFC

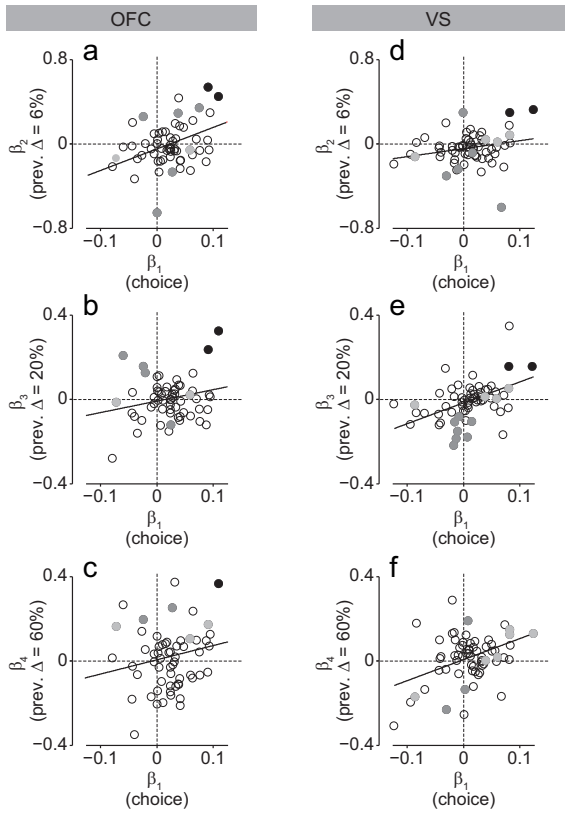

Figure 6.7. Average firing rate GLM parameters for current choice and previous trial outcome interacting with stimulus difficulty. **(a,b,c)** Previous outcome estimators for difficult choices (a), medium difficulty (b) and easy choices (c) as a function of choice estimator β_1 for OFC neural data. **(d,e,f)**. Same as (a,b,c) but for VS data. Black indicate cells that show significance ($p < 0.05$) for both plotted estimators, light gray for choice components and dark gray for previous outcome components. Dark lines represent best linear regressions for the plotted data.

Estimators	ρ	p
β_1,β_2	0.200	0.11
β_1,β_3	0.237	0.06
β_1,β_4	0.167	0.19

Table 6.2. Spearman's rank correlation between GLM estimators for VS data.

showed a property that resembles the choice bias updating curves (Figure 6.8a). They indicate that the previous outcome had a larger influence in neural activity when the stimulus was difficult (6% contrast). For the VS, these slopes show that the change is not dependent of difficulty as much as they depend of outcome (Figure 6.8b).

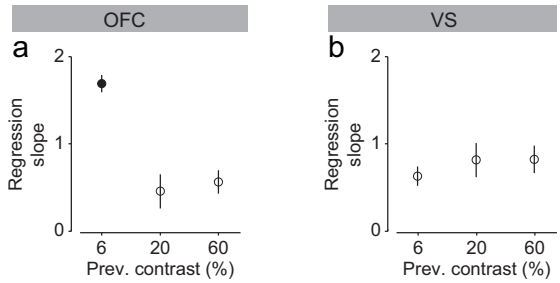

Figure 6.8. Slope of linear regressions for GLM difficulty parameters. **(a,b)** Best fit slopes of the linear regressions of Figure 6.7 plotted as a function of previous contrast for both OFC (a) and VS (b) data. Filled point indicate significance in Spearmann's rank correlation (Tables 6.1 and 6.2).

6.4.3 Example cells

Visual inspections of OFC #39 (Figure 6.9) shows that neuronal activity is greatly modulated by the previous outcome, if that same trial was of a difficult kind (green lines). However, despite overall activity going up, the dynamics revealed to be rich as, for example, there is an inversion of

sign when the rat leaves the decision poke. Comparatively, easy trials did not modulate as much the mean response as it coarsely overlaps with the mean PETH (grey lines).

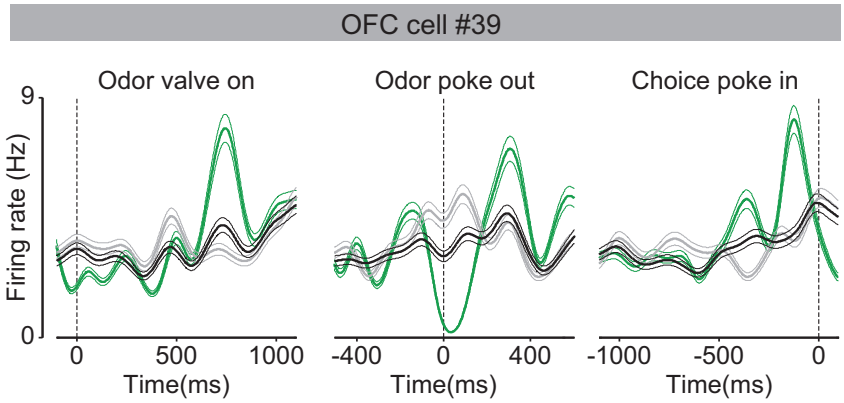

Figure 6.9. PETHs for example OFC cell modulated by previous outcome. Three plots are presented with alignment to three different events: odor valve on (beginning of stimulus presentation), odor poke out and choice poke in. Green lines represent PETH done in trials that followed a 6% contrast left choice that was rewarded. Gray lines represent trials that followed a 60% contrast left side reward. Black lines depict the average PETH. All lines are mean ± SEM over trials.

VS cell #39 also demonstrated an increase of firing rate after a difficult rewarded decision (Figure 6.10). This increase was specially significant in the moments immediately following entrance at the choice poke. Conversely, easy trials showed no modulation of activity when compared with a difficult trial.

Taking these results into account, we concluded that neural activity at both OFC and VS show modulation of previous outcomes in current choices. Additionally, OFC activity showed interaction between outcomes and difficulty, a result that is expected from an evidence accumulating particle, according to RL-DDM. VS, on the other hand, did not show this

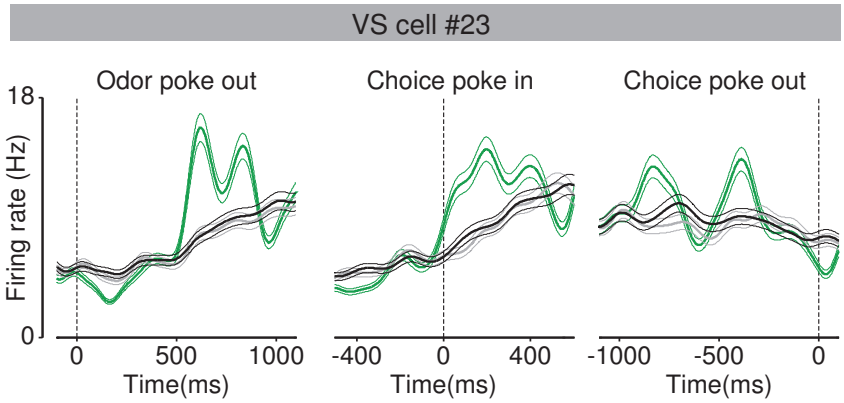

Figure 6.10. PETHs for example VS cell modulated by previous outcome. Three plots are presented with alignment to three different events: odor poke out, choice poke in and choice poke out (end of trial). Green lines represent PETH done in trials that followed a 6% contrast left choice that was rewarded. Gray lines represent trials that followed a 60% contrast left side reward. Black lines depict the average PETH. All lines are mean ± SEM over trials.

modulation at a population level, as only 3% of the recorded cells showed modulation of difficulty. These results and possible explanations/issues for the observed differences are explored in more detail in the next Chapter.

Chapter 7

Discussion

"Did I ever tell you what the definition of insanity is? Insanity is doing the exact... same fucking thing... over and over again expecting... shit to change... That. Is. Crazy."

– Vass, *Far Cry 3*

7.1 Speed-accuracy tradeoffs depend on the nature of the task at hand.

Our results demonstrate that rats show different speed-accuracy tradeoffs (SAT) depending on the task at hand. When challenged to identify odors at low concentrations, rats show a significant increase of reaction time that is accompanied by performance degradation (Figure 2.4c,d). In contrast, when the challenge is to categorize mixtures of two odors in different proportions, rats show only a small increase in reaction time (Figure 2.4e,f).

In most perceptual decision-making tasks (PDM), errors are assumed to arise from uncertainty about the stimulus, or noise in the respective sensory system. This implies that performance typically increases with stimulus sampling duration (Brunton et al., 2013; Kiani & Shadlen, 2009). However, there are situations, such as the odor mixture categorization task presented in this thesis and described before (Uchida & Mainen, 2003; Zariwala et al., 2013), where SAT is not observed, suggesting that the rapid performance observed in this task is not simply a tradeoff of accuracy for speed. In fact, a single sniff is all a rat needs to perform at maximum level for the categorization task (Uchida & Mainen, 2003).

We also controlled for changes in total concentration of mixtures (Figure 2.2). We saw scaling of the same effect within each concentration level, i.e. total concentration modulated reaction time greatly, but, within same total concentration, mixture contrasts did not (Figure 2.5c). Conversely, mixture contrast was still degrading performance (Figure 2.5b). Does this imply that SATs are not existent in mixture categorization? This is not necessarily the case as other variables might modulate performance and RTs, such as reward magnitude, reward rate and training (Zariwala et al., 2013). However, considering our results, it is clear that the currency used between speed and accuracy is intrinsically different and modulated by the

nature of the question being asked. Even when considering very similar conditions that only differ slightly in the stimulus being presented, such as the case of odor identification versus mixture categorization. Taken together, these behavioral results not only help conciliate contradicting observations from the field of rodent olfactory PDM (Rinberg et al., 2006b; Abraham et al., 2004; Uchida & Mainen, 2003) but also bring forward a framework for the understanding of the computational requirements behind such decisions.

7.2 Perceptual decision-making is driven by sensory uncertainty...

We used a standard drift-diffusion model (DDM) to test the role of stimulus uncertainty in both identification and categorization tasks. We found that the conventional form of DDM was sufficient to explain both tasks separately (Figure 3.4). However, the parsimonious hypothesis in which both tasks share the same stimulus input parameters revealed unfeasible (Figure 3.6). These results show that the task difference cannot be explained by merely stimulus noise (Chapter 3), even with the addition of reward-dependent choice biases (Figure 5.2).

The DDM parameters obtained from the identification task's fit (Table 3.2) predict that the relative amount of evidence for the categorization task should be extremely high, even for low contrast mixtures (Figure 3.5, black lines). Yet, behavioral accuracy in this task is still degraded. It is known that signal-to-noise ratio can be increased through temporal integration (Gold & Shadlen, 2007). However, this effect is only possible if noise is not temporally correlated over time. For instance, correlated noise in the activity across a neuronal population can dramatically limit the usefulness of pooling spikes across more neurons in order to increase the signal-to-noise ratio (Zohary, Shadlen, & Newsome, 1994). Addition-

ally, noise correlations limit the ability of averaging noise through repeated sampling (Uchida et al., 2006). Considering these facts, one is led to the conclusion that the usefulness of temporal integration depends on the nature of the limiting noise. Therefore, the differences seen in the two independently fitted DDMs (Table 3.3) must be due to structurally different types of noise.

7.3 . . . but also by category uncertainty.

Previous work had suggested that the categorization task was limited by uncertainty related to the categorical separation between odor A and B (Kepecs, Uchida, Zariwala, & Mainen, 2008; Kepecs, Uchida, & Mainen, 2008). In other words, it was proposed that the inherent challenge for the rats was not detecting which stimulus was presented (Uchida & Mainen, 2003), but in fact correctly recalling a separation bound that is drifting in a trial-by-trial fashion (Kepecs, Uchida, & Mainen, 2008). We further validated this proposal by adding random variability to our DDM model, i.e. by adding random "memory noise" (Figure 3.7 and 3.8).

The addition of trial-by-trial variability to the DDM implies an additional source of uncertainty. However, in its most basic form, this variability does not imply a source of noise that is correlated. In fact, the effects of random variability in drift-rate can be mitigated by prolonged integration (Ratcliff & Smith, 2004). Considering that rats do not show improved performance in the categorization task even when instructed to sample for longer (Zariwala et al., 2013), we specifically hypothesized that this type of uncertainty must have a trial-by-trial structural dependency.

We proposed that odor mixture categorization is limited by constant updating of the category boundary between left and right set by the experimenter and that must be learned by the subjects through trial-by-trial reinforcement (Zariwala et al., 2013). Indeed, by looking at the influence

of trial history on the choice of the animals, we observed a clear trial-by-trial updating of the animal's choice function, which depended both on the difficulty of previous trial and outcome (Figure 4.4).

Therefore, if uncertainty about the precise category boundary dominates over stimulus uncertainty, the benefits of integration within a single trial would be curtailed. In this regard, tasks that are dominated by uncorrelated sensory noise may indeed show the expected benefits of extended stimulus sampling (Roitman & Shadlen, 2002; Brunton et al., 2013). Based on our behavioral and modeling results, we believe that the identification task falls on this category, while mixture categorization does not.

7.4 Category uncertainty does not imply noise, just a bad strategy.

We expanded our DDM by introducing a reinforcement learning process, the kind theorized to drive stimulus-response learning (RL-DDM). With the combination of these three factors – stimulus noise, reward bias and categorical boundary learning – the RL-DDM model did not only fit the average performance data (Figure 5.7c,d), but also predicted the choice biases on the recent history of stimuli, choices and rewards (Figure 5.7e,f). Furthermore, the RL-DDM model was able to predict performance over an interpolated stimulus space combining both tasks in the same sessions (Figure 2.5), ruling out differences in strategies between the two tasks and arguing that rats used the same decision-making system while detecting and categorizing odors.

We found that mixture categorization performance is more susceptible to category boundary fluctuations than odor identification (Figure 5.11). This is due to high stimulus input that always exists in this task. RL rules dictate that whenever there is a mismatch between the predicted

and observed outcome, a change in weights will occur. However, the multiplication implies amplification when sensory evidence is large (Equation 5.12).

Our model is in agreement with previous observations for the categorization task (Uchida & Mainen, 2003; Zariwala et al., 2013). In particular, the much smaller tradeoff between accuracy and reaction times observed in this task is not due to a lack of quality of stimulus input. In fact, our model indicates that stimulus strength is extremely high for the categorization task (Figure 5.11b). This would suggest that this type of uncertainty is correlated between trials and that extended integration would not favor a better decision.

To test this we ran an additional RL-DDM simulation (Figure 7.1). We saw that performance remains unaltered in mixture categorization with a two-fold increase in the diffusion threshold θ of our RL-DDM, contrasting with what would be predicted for odor identification. This is in agreement with the observation that one sniff is enough for maximum performance in mixture categorization (Uchida & Mainen, 2003). Weight fluctuations, which impair performance in a trial-by-trial basis, cannot be easily filtered out within the integration process (and only if weight multiplication occurs within the diffusion process). On the other hand, identification task is highly driven by stimulus noise, which is reflected within the diffusion process, and thus favored by extended integration in order to make better decisions. We thus conclude that the observation of different speed-accuracy tradeoffs is just due to different computational requirements in the two tasks.

We have demonstrated that the simple scenario of detecting a noisy stimulus is insufficient to capture all the details occurring in a two-forced choice task. Two other effects have to be incorporated in order to explain the differences observed here. First, the effect of bias, as previously described (Busse et al., 2011), and second, a novel result of on-going

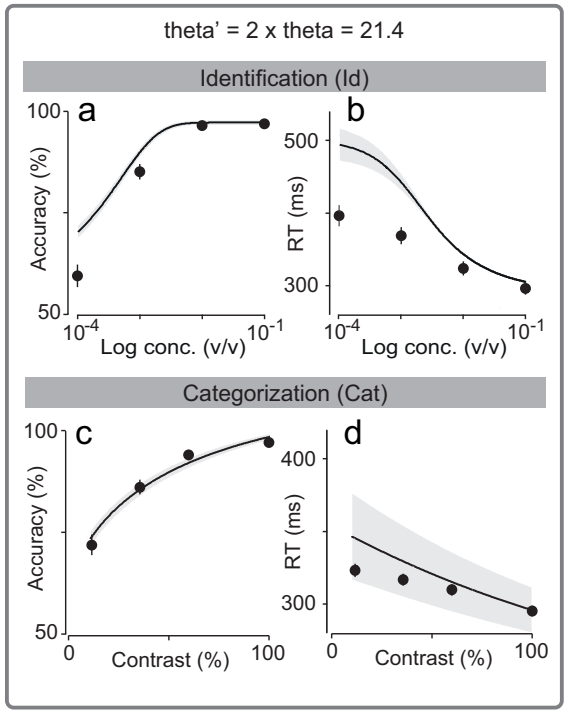

Figure 7.1. RL-DDM predictions with threshold increase for both tasks. **(a,b)** Psycho- and chronometric predictions for identification task considering a two-fold increase in threshold for RL-DDM. **(c,d)** Same as (a,b) but for categorization task. Error bars show SEM across rats and trials. Lines show predictions from the RL-DDM. Grey shaded area shows the models' predicted SEM for the same number of trials as seen in the data.

stimulus-dependent RL (Figure 4.4). These results show that, given the right conditions, learning can be detrimental for the rats' performance while categorizing stimuli.

Nevertheless, these trial-by-trial dependencies might not be observed in all tasks. For example, in the case of visual discrimination of random dots coherence (Roitman & Shadlen, 2002; Palmer et al., 2005) or auditory discrimination of clicks (Brunton et al., 2013) the stimulus being presented

is already lateralized, i.e. they inherently signal the correct decision. We hypothesize that these tasks will show reduced updating effects, as the 0 coherence boundary represents a strong prior that the subjects have experienced extensively, contrasting with the 50/50 odor mixture separation boundary, which represents an arbitrary mapping to left/right responses that must be learned.

We have shown ongoing learning as compromising the rats' ability to categorize odors. We have ensured that ongoing learning is not due to incomplete learning as the rats present stable performance over the analyzed data (Figure 4.1, 4.2 and 4.3). This suggests some other interfering factor or some deliberate strategy of the rats. One possible interpretation is that the rats' performance is limited by the inability to keep track of very long trial histories. An ideal decision maker would learn to set the perfect decision boundary by averaging over all the trials that is has been exposed to. Imperfect memory would imply that the most recent trials will have a larger effect in deciding what to do with a given stimuli (conditional on ongoing learning occurring, $\alpha \neq 0$, i.e. that the most recent trial is still affecting performance).

An ideal observer would turn off the effect of on-going stimulus learning in order to maximize its rewards after the contingencies have been learned (Summerfield & Tsetsos, 2012). However, in this particular case the observer's belief that the environment is ever-changing indicates that an optimal strategy is not being deployed. We have shown that the additional source of uncertainty in categorization task is due to a deterministic rule that amplifies stimulus noise and generates more errors.

7.5 The neural circuitry of olfactory decision-making.

Our results are consistent with the idea that identifying an odor at low concentrations and distinguishing closely related mixtures of odorants call for very different computational demands. More specifically, we hypothesized that, in the case of the identification task, the observed accuracy and RTs are due to the reduced amount of odor information at lower concentrations.

The neural circuitry behind odor detection is extremely difficult to define (Vicente, 2015). This is mainly due to a lack of electrophysiological data available from awake and behaving animals for changing concentrations. Most of the available datasets are from anesthetized animals whose neural responses are very different from awake animals (Rinberg, Koulakov, & Gelperin, 2006a). Nevertheless, recordings from olfactory receptor neurons (ORNs) and the olfactory bulb (OB) have shown that increasing odor concentration leads to an increase in the number of recruited ORNs and glomeruli and in response amplitude (Haberly, 2001; Carey, Verhagen, Wesson, Pirez, & Wachowiak, 2009; Spors & Grinvald, 2002; Johnson & Leon, 2000). Conversely, odor concentration also modulates neurons' response latencies (Rinberg et al., 2006a; Reisert & Matthews, 2000, 2001; Duchamp-Viret, 1999; Cang & Isaacson, 2003). Based on these observations, a variety of schemes have been proposed for spike encoding of odor intensity, namely a spike rate code and a spike latency code. However, there is still not a unified view of how odor intensity is encoded in the brain (Cury & Uchida, 2010; Bathellier, Buhl, Accolla, & Carleton, 2008). For instance, even though it has been shown that response latency decreases with increasing concentration, the range of magnitudes observed is remarkably distinct, e.g. from $\sim$50 ms latency for a 10-fold change in mouse isolated ORNs (Reisert & Matthews, 2001) to >900 ms

change for a 150-fold difference in anesthetized rat ORNs (Duchamp-Viret, 1999), or even examples of mitral cells (MCs) that show increased firing latencies with increasing concentration (Bathellier et al., 2008). Moreover, odor-evoked electro-olfactogram (EOG) recordings from the ORNs of anesthetized rats have shown that the EOG onset latency is almost unaltered across concentrations (Duchamp-Viret, 1999).

Regarding our results on the odor identification task, we cannot rule out the existence of sensory delays, the examples that we cited indicate that there are heterogeneous response profiles across the odor-responsive cells. Nevertheless, the existence of neurons with minimally concentration-dependent delays would be sufficient for animals to begin to perform odor identification rapidly.

In our RL-DDM, framework we hypothesize that the momentary evidence that arises from the sensory layer is probably related to neural activity represented in the ORNs and MCs. Downstream from this area, we believe that PIR, a highly associational sensory cortex (Stettler & Axel, 2009), is integrating and properly identifying the mixture of odorants being presented to the animal. Considering the outperformance from a simple DDM in categorization task (Figure 3.5), we predicted that PIR mixture representations should be in fact better than the behavior demonstrated by the rats. These predictions are in agreement with a decoding process that outperforms the behavioral data in the mixture categorization task (Miura et al., 2012). Other sensory areas have shown similar results. For example, an optimal decoder of neural activity in the primary visual cortex V1 during a detection task was shown to outperform monkey's behavior both in speed and accuracy (Chen, Geisler, & Seidemann, 2008). Altogether, these results originally implied that there might be sources of noise downstream to primary structures such as V1 or anterior PIR, that limit behavioral performance, both in terms of RTs and accu-

racy. Here, we reject that assumption and hypothesize that in fact an RL-driven suboptimal decoder is miscategorizing these responses.

When it comes to integration and decision variables, a number of different regions have been proposed as key in the process of PDM. In particular, the pariental cortex has long been implicated to be involved in the transform from stimulus to action (Gottlieb, 2007). The lateral intraparietal area (LIP) exhibit firing rates that accelerate with the level of evidence (in the form of random dot motion coherence) of a stimulus (Roitman & Shadlen, 2002). Additionally, microstimulation of LIP showed that responses were biased towards the respective stimulated receptive fields (Hanks et al., 2006). In rodents, the posterior parietal cortex (PPC) is believed to be playing an analogous role to primates' LIP (Hanks et al., 2015). In particular, Hanks et al. showed that while PPC encoded graded value of the accumulating evidence, the frontal orienting fields (FOF) had a more categorical encoding that indicated, throughout the trial, the decision provisionally favored by the evidence accumulated so far. Despite the difference in modality (in this case, audition) we hypothesize that similar results should be observable in our olfactory tasks. An important question regarding the transform from stimulus into values and then into actions is to pin point where weight updating occurs. We are unclear of where this multiplication of weights and stimulus information occurs, if prior or post accumulation. Electrophysiological recordings from PPC and FOF should help clarify these issues and fine tune our hypothesized model.

Due to the significant importance of RL in EDM, an important area to consider in our model implementation is the OFC. OFC has been shown to play a significant part in value monitoring (Kable & Glimcher, 2007; Padoa-Schioppa & Assad, 2006, 2008; Padoa-Schioppa & Cai, 2011; Schoenbaum et al., 2011). In particular, for the case of olfaction, OFC has been implicated in the integration of sensory evidence (Bowman et al., 2012); however, the authors could not rule out the hypothesis that the

observed OFC activity could had been related with a confidence signal (Kepecs, Uchida, Zariwala, & Mainen, 2008). Kepecs, Uchida, Zariwala, and Mainen showed that OFC activity reflected the level of confidence of a rat during the categorization task. Additionally, OFC inactivation affected the confidence report of rats while not affecting the behavioral performance in the task (Lak et al., 2014).

In our RL-DDM, confidence is related to the amount of time that has passed - a long integration process reflects a larger bound collapse implying a significant prediction error in Equation 5.12. An alternative frame of mind is to consider two competing races, confidence in a given decision can then be computed by calculating the relative difference between the two options (red line in Figure 5.13a). Regardless, if confidence is coded in the OFC, then the activity of these neurons should reflect the predicted changes in weights from an RL algorithm.

We showed that this was indeed the case (Figure 6.7). More importantly, we demonstrated that OFC activity was modulated by the quality of the previous stimulus presented, with more difficult stimuli presenting a larger change in activity (Figure 6.8). These results largely support the idea of ongoing updating of stimulus value on a trial-by-trial basis.

VS was previously implicated in inferring value information necessary to drive reward-based decisions (Bissonette et al., 2013; Cromwell & Schultz, 2003; FitzGerald, Schwartenbeck, & Dolan, 2014; Haber & Knutson, 2010; Ito & Doya, 2009; van der Meer & Redish, 2011, 2009; Roesch, Singh, Brown, Mullins, & Schoenbaum, 2009) and is suggested to play the role of critic, in model-based reinforcement learning, by monitoring decisions and computing outcome expectations (Mannella, Gurney, & Baldassarre, 2013; Ito & Doya, 2011; van der Meer & Redish, 2011). The VS role in modulating goal-directed behaviors is normally seen as being shared with OFC (Botvinick et al., 2009; Hare et al., 2008; McDannald et al., 2011). In the context of the mixture categorization task, VS

also presented decision uncertainty/confidence correlated activity (Costa, 2015).

We found that VS also showed modulation of activity dependent on the outcome of the previous trial (Figure 6.4). However, stimulus difficulty did not show a significant effect (Figure 6.8). The differences seen from OFC to VS were surprising but also intriguing. An important difference between the two areas is related with the timing of the events. OFC activity was more significant during sampling and movement time, while VS showed confidence related signals during movement and reward anticipation (Costa, 2015). A possible interpretation is that OFC is essential to coordinate the "global value currency" from stimulus to actions and that that transform is not blind to the stimulus quality. Conversely, the VS could be coding action values and reflecting trial-by-trial modulations that depend of choice and reward, but blind to stimulus quality. These results would imply analogous properties to what was seen in the interplay between PCP and FOF (Hanks et al., 2015).

Interestingly, Costa and Mainen showed that the rats behavioral report (time willing to wait for a reward) was unaffected by reward magnitude, despite significant changes in choice bias (Figure 6.2; Costa, 2015). A computational model inspired by standard SDT was proposed to generate such responses. The model goes as follows: each trial, a sample from a Gaussian distribution is made for the stimulus representation, s_i; this sample is then compared to a noisy bound, B, also sampled from a Gaussian distribution; a choice is determined by the sign of the difference $(s_i - B)$. In rewarded blocks the bound is influenced by a constant ζ, which is unrelated to the stimulus but used to compute choice as $sign(s_i - B \pm \zeta)$. In order to generate the confidence report of the rats, Costa and Mainen defined two types of confidence: an "un-biased confidence" that is exclusively related with stimulus and computed considering a function that transforms $(s_i - B)$; and a "biased confidence" that does the same but takes into ac-

count ($s_i - B \pm \zeta$). Non surprisingly, they were able to generate responses that matched the behavioral data. However, not so expected was the fact that confidence neural signals of OFC and VS matched this difference in confidence "types". OFC population coded *un-biased* confidence, while VS showed *biased* confidence signals.

The RL-DDM considers two types of updating effects. One is general and occurs for any stimulus. We hypothesized that this effect is related to all non-related task inputs, in a similar fashion as what was done by Busse et al. (2011). This general updating is extremely important for the identification task when stimulus input is extremely low (Figure 5.7e). Its effect on the integration process is to offset the starting point of integration though w_b. The other type of updating is stimulus dependent and varies with the quality of the previous stimulus. Conversely, this type of updating is extremely important for the categorization task (Figure 5.7f), and is represented in our RL-DDM as variability in drift rates on a trial-by-trial manner. This model implies that two processes are occurring simultaneously: matching of reward due to all non-olfactory related processes, and learning regarding the olfactory stimulus contingencies. Are these computations then occurring at different brain areas?

If OFC is considered as an "unbiased confidence" estimator, then this would be reflected as calculating the difference between evidences from RL-DDM, i.e. $(w_1 s_1 - w_2 s_2)/\theta_0$ in our formulation. On the other hand, if VS is computing "biased confidence" from RL-DDM that would be reflected as the starting position (bias) of DDM and thus the value of w_b. These assumptions could then explain the differences seen for OFC dependency on previous stimulus, and VS dependency on just reward outcome. If they are correct, then an important point of investigation is to find the downstream area in which OFC and VS converge and merge signals. A potential candidate is the pre-motor cortex which has been shown to

be responsible for "leaving decisions" in a waiting time task (Murakami, Vicente, Costa, & Mainen, 2014).

However, even though we were able to recognize responses that matched our predicted dynamics from the RL-DDM, this represented a reduced number of the whole population for both OFC and VS. Part of the reason why this might occur is due to the fact that a significant fraction of neurons in OFC and VS compute the quality of a decision in regards to a bound, rather than the value of a particular decision poke (Kepecs, Uchida, Zariwala, & Mainen, 2008; Kepecs, Uchida, & Mainen, 2008), a fundamental assumption that was incorporated in our GLM (Equations 6.3 and 6.4). Incorporating confidence signals and changing that same assumption might show a richer dataset to explore. Additionally, our GLM analysis was done on the complete course of the trial (Equation 6.1). As the information of each area seems to be significant at different time points (Costa, 2015), a route to pursue is to identify these moments and address whether a finer resolution shows RL dependent modulation of neural activity. These effects require further analysis of our neuronal dataset.

7.6 Future directions.

In the previous section we focused on the results from recorded areas OFC and VS (Costa, 2015). We also hypothesized that the involvement of both PPC and FOF should be of significant importance in this task. as they have been previously shown to participate in an auditory accumulation of evidence task (Hanks et al., 2015). Electrophysiological recordings from these areas, from rats performing both identification and categorization tasks, would significantly ameliorate our understanding of olfactory PDM, and fine-tune the parameter space of our RL-DDM.

However, these areas are clearly involved after some sensory processing steps have been resolved. Considering that the main difference between the two tasks resides in the nature of the presented stimuli, a fundamental future step would be to identify and record from the first stages involved in olfactory processing.

There is evidence that odor representations suffer a profound transformation from OB (Cury & Uchida, 2010; Shusterman, Smear, Koulakov, & Rinberg, 2011) to the anterior PIR (Miura et al., 2012). Namely, OB information is conveyed within the first 100 ms after inhalation onset at a resolution of tens of milliseconds (Cury & Uchida, 2010; Shusterman et al., 2011), while in the aPIR reliable odor information is provided by total spike counts over the entire sniff cycle (Miura et al., 2012). Given these temporal differences in the OB and aPIR, namely in terms of the type of information relevant for each area, simultaneous recordings from these two areas, in rats performing the odor identification task, would provide important insights about how odor intensity is encoded in the brain and how this information is used in a decision-making process.

We showed that the mixture categorization task is limited by trial-to-trial fluctuations introduced by RL signals. This influence of trial history on odor perception should be most likely reflected in the neural activity of a brain structure that receives coincident input from the olfactory and reward systems. A potential candidate is the olfactory tubercle (OT). The OT receives monosynaptic olfactory input from both the OB and the PIR (Wesson & Wilson, 2011) and displays odorant-evoked responses (Wesson & Wilson, 2011; Murakami, Kashiwadani, Kirino, & Mori, 2005). Additionaly, the OT is considered part of the ventral striatum and is heavily interconnected with the reward system (Ikemoto, 2007), receiving projections from several areas, including the ventral tegmental area (VTA), the nucleus accumbens and the substantia nigra (Wesson & Wilson, 2011). We hypothesize that the OT is playing a significant role in the interface

between olfactory information into actions. The modulation of OT activity due to RL signals could be reflected in the variability of OT neuronal responses, as observed for single dorsal premotor neurons in an arm reach countermanding task (Marcos et al., 2013), where the variability (and not the mean) of neural activity in a given trial increased with the number of previous trials containing a stop signal. We predict that signals from the OT in a mixture categorization task will present mean activity modulation dependent on the interaction between the quality of the previous trial and presence (or absence) of a reward.

How would these updating effects occur at the OT level? We speculate that the modulation could be mediated by neuromodulatory inputs, in particular by dopamine (Schultz, 2015). An interesting approach would then be to use optogenetic tools in order to test and manipulate the feedback signals from the VTA. Additionally, serotonin (5-HT) from the raphe nucleus presents itself as an interesting candidate, as it has been shown to influence OT activity in rats (Wesson & Wilson, 2011). Preliminary work from our laboratory has demonstrated that 5-HT neurons from the dorsal raphe nucleus (DRN) respond to reward-predictive cues in a way similar to a prediction error (Matias, Lottem, Dugué, & Mainen, 2015) and that optogenetic activation of DRN 5-HT neurons, in anesthetized rats, produces a rapid and profound inhibition of spontaneous (but not odor-evoked) firing of olfactory cortex neurons, which was multiplicative and frequency-dependent (Lottem, Lőrincz, & Mainen, 2015). Slice experiments also revealed that DRN 5-HT activation inhibits cortical feedback compared to feedforward input (Lottem et al., 2015). We postulate that the same type of modulation that was observed in PIR could also happen in the OT, which receives direct input from the raphe nucleus (Wesson & Wilson, 2011).

Overall, these hints given by the olfactory neural circuits clearly indicate that recordings from the PC and OT in rats performing both the

identification and categorization tasks would provide valuable information about the strategies employed to solve these tasks. Conversely, the usage of optogenetic tools could then help fine-tune the role of other areas in PDM, such as areas from the reward system.

Lastly, a relevant question to the field of olfaction regards the impact of sniffing on sensory processing. When performing a mixture categorization task, similar to the one used in our study, rats sample odors at ∼8 Hz (Uchida & Mainen, 2003), and did not change as task difficulty is increased by lowering odorant concentration (Wesson, Verhagen, & Wachowiak, 2009). Sniffing has not been measured for the identification task, but if we consider the difference in RTs for lower concentrations, a sniffing at ∼8 Hz and accept that sniff frequency is not changing across concentrations, then rats should be taking an extra 1-2 sniffs for lower concentrations. However, if that is the case, what happens in between sniffs? Sniffing has been shown to provide a reference frame for neural responses in the OB and PIR (Miura et al., 2012; Cury & Uchida, 2010; Shusterman et al., 2011) and the timing of MCs firing in the OB is conserved across various respiration frequencies (Cury & Uchida, 2010). Nonetheless, it is not known how this information is used to build a representation of evidence, and what is the influence of extra sniffs. It is unclear whether each extra sniff would be integrated over time or considered as an independent sample of evidence. Recent evidence has shown that odor representation in the OB evolves after the first breath and persists as an odor afterimage (Patterson, Lagier, & Carleton, 2013). The way each one of these bits of information is handed to OT and OFC would be critical to understand the mechanisms underlying the decision process across multiple sniffs. Simultaneous sniffing recordings in the identification task would help shed light on these questions (achievable through the usage of a thermocouple nose implant; Uchida & Mainen, 2003).

7.7 Final remarks.

As a last point for discussion, we would like to highlight an important contribution of our work for the field of PDM. Concepts like ideal observers and optimality are delicate when one considers the naturalistic and ever-changing environment that has played a significant role in evolution. That changing environment would imply that ongoing learning is actually an optimal strategy. The psychophysics-like experimental paradigm is a highly artificial one in the sense that outcomes and states are crystallized. However, it is unlikely that this would be the case in a more naturalistic environment, where, due to environmental dynamics, odors could signal different outcomes, rewards and states over time.

The role of detection is easily understandable from a perspective of nothing against something, such as in the case of odor identification. However, the separation bound between olfactory (or any perceptual modality for that matter) objects is hard to believe as fixed and unchanging in a naturalistic environment. For example, distinguishing between an un-ripe, ripe or spoiled *"green-blue fruit"* can be the difference between living or dying. But these states are probably dependent of multiple environmental factors that the brain has to integrate and rapidly adapt to. In fact, olfaction has been shown to depend highly of context in the evaluation of odor pleasantness (Leleu et al., 2015).

A normal, ever-changing, environment would imply adaptability and never ending learning as the optimal strategy. The results presented in this thesis are consistent with a recent proposal that suboptimal inference, as opposed to internal noise, is a major cause of behavioral variability (Beck et al., 2012). In this case, the suboptimal inference is the result of assuming that the world is dynamic when, in fact, it is static.

But is that assumption unfair? Is the superstitious behavior of our rats due to their own *"stupidity"* or our own as experimentalists? An al-

ternative explanations is that evolution has made extremely advantageous (and thus hardwired) to develop rapidly adaptive strategies. These type of studies in PDM are fundamental for our understanding of neural dynamics. Through them we could unveil the functional evolutionary pressures that the NS has faced throughout the ages and thus design better and more significant tasks for our understanding of the brain. All in all, it is clear that the question is not so much whether there is an optimal or sub-optimal strategy for a given set of artificial conditions, but in the direction of understanding why that gap exists.

As a side note, let me narrate you the events of one particular session with a rat that I kept in my memory. In that session, the rat rapidly evolved into an incredibly biased state, with only right side choices being signaled. After $\sim$100 trials, I decided to interrupt the task as clearly something was very wrong. After debugging the behavioral set-up, I realized that the water delivery system was not working correctly and not delivering rewards on the left side. To make matters worse, it was delivering rewards to the right side always (after initializing the trial with a center poke). This adaptability from the rat allowed him to rapidly change its bias, ignore the stimulus being presented, and, thus, exploit the experiment's (and experimenter's) flaw.

One should not consider that there is a fundamental problem in the way a task is designed, or if subjects are not learning it properly. A state of constant flux, combined with plasticity and ongoing modulation of responses, have to be key and fundamental properties of the brain. As scientists, let us not forget and run away from this evidence.

A lack of optimality is not a sign of an error in either the experimenter or the behavioral subject. It's, in fact, an opportunity to answer a driving and important question: "That was unexpected... but, *why*?"

References

Abby, S., & Daubin, V. (2007). Comparative genomics and the evolution of prokaryotes. *Trends Microbiol.*, *15*(3), 135–141. doi: 10.1016/j.tim.2007.01.007

Abraham, N. M., Spors, H., Carleton, A., Margrie, T. W., Kuner, T., & Schaefer, A. T. (2004). Maintaining accuracy at the expense of speed: stimulus similarity defines odor discrimination time in mice. *Neuron*, *44*(5), 865–76. doi: 10.1016/j.neuron.2004.11.017

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychol. Rev.*, *98*(3), 409–429. doi: 10.1037/0033-295X.98.3.409

Ash, R. B. (2012). *Information Theory*. Dover Publications.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *J. Exp. Psychol. Learn. Mem. Cogn.*, *14*(1), 33–53.

Balleine, B., & Dickinson, A. (1998). the Role of Incentive Learning in Instrumental Outcome Revaluation By Sensory-Specific Satiety. *Learn. Behav.*, *26*(1), 46–59.

Bathellier, B., Buhl, D. L., Accolla, R., & Carleton, A. (2008). Dynamic ensemble odor coding in the mammalian olfactory bulb: sensory information at different timescales. *Neuron*, *57*(4), 586–98. doi: 10.1016/j.neuron.2008.02.011

Baum, W. M. (1974). On two types of deviation from the matching law: bias and undermatching. *J. Exp. Anal. Behav.*, *22*(1), 231–242. doi: 10.1901/jeab.1974.22-231

Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *J. Exp. Anal. Behav.*, *32*(2), 269–281. doi: 10.1901/jeab.1979.32-269

Baum, W. M. (1983). Matching, statistics, and common sense. *J. Exp. Anal. Behav.*, *39*(3), 499–501. doi: 10.1901/jeab.1983.39-499

Baum, W. M., & Davison, M. (2004). Choice in a variable environment: visit patterns in the dynamics of choice. *J. Exp. Anal. Behav.*, *81*(1), 85–127. doi: 10.1901/jeab.2004.81-85

Beck, J., Ma, W., Kiani, R., Hanks, T., Churchland, A., Roitman, J., ... Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, *60*(6), 1142–1152. doi: 10.1016/j.neuron.2008.09.021

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not Noisy, Just Wrong: The Role of Suboptimal Inference in Behavioral Variability. *Neuron*, *74*(1), 30–39. doi: 10.1016/j.neuron.2012.03.016

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.*, *10*(9), 1214–1221. doi: 10.1038/nn1954

Bengson, J. J., Kelley, T. A., Zhang, X., Wang, J.-L., & Mangun, G. R. (2014). Spontaneous neural fluctuations predict decisions to attend. *J. Cogn. Neurosci.*, *26*(11), 2578–2584. doi: 10.1162/jocn_a_00650

Bennett, D. J. (2009). *Randomness*. Harvard University Press.

Bissonette, G. B., Burton, A. C., Gentry, R. N., Goldstein, B. L., Hearn, T. N., Barnett, B. R., ... Roesch, M. R. (2013). Separate populations of neurons in ventral striatum encode value and motivation. *PLoS One*, *8*(5), e64673. doi: 10.1371/journal.pone.0064673

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.*, *113*(4), 700–765. doi: 10.1037/0033-295X.113.4.700

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement-learning perspective. *Cognition*, *113*(3), 262–280. doi: 10.1016/j.cognition.2008.08.011

Bowman, N. E., Kording, K. P., & Gottfried, J. A. (2012). *Temporal Integration of Olfactory Perceptual Evidence in Human Orbitofrontal Cortex* (Vol. 75) (No. 5). doi: 10.1016/j.neuron.2012.06.035

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993).

Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.*, *10*(6), 1157–1169.

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. *Science (80-.).*, *340*(6128), 95–98. doi: 10.1126/science.1233912

Burke, K. A., Franz, T. M., Miller, D. N., & Schoenbaum, G. (2008). The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards. *Nature*, *454*(7202), 340–344. doi: 10.1038/nature06993

Busse, L., Ayaz, A., Dhruv, N. T., Katzner, S., Saleem, A. B., Scholvinck, M. L., ... Carandini, M. (2011). The Detection of Visual Contrast in the Behaving Mouse. *J. Neurosci.*, *31*(31), 11351–11361. doi: 10.1523/JNEUROSCI.6689-10.2011

Buzsáki, G., Peyrache, A., & Kubie, J. (2015). Emergence of Cognition from Action. *Cold Spring Harb. Symp. Quant. Biol.*, *79*, sqb.2014.79.024679–. doi: 10.1101/sqb.2014.79.024679

Cang, J., & Isaacson, J. S. (2003). In vivo whole-cell recording of odor-evoked synaptic transmission in the rat olfactory bulb. *J. Neurosci.*, *23*(10), 4108–4116.

Carey, R. M., Verhagen, J. V., Wesson, D. W., Pirez, N., & Wachowiak, M. (2009). Temporal structure of receptor neuron input to the olfactory bulb imaged in behaving rats. *J. Neurophysiol.*, *101*(2), 1073–1088. doi: 10.1152/jn.90902.2008

Chalmers, A. F. (1999). *What is this Thing Called Science?* University of Queensland Press.

Changeux, J.-P., & Dehaene, S. (2008). The Neuronal Workspace Model: conscious processing and learning. *Learn. Theory Behav.*, *1*, 729–758. doi: http://dx.doi.org/10.1016/B978-012370509-9.00078-4

Chen, Y., Geisler, W. S., & Seidemann, E. (2008). Optimal temporal decoding of neural population responses in a reaction-time visual detection task. *J. Neurophysiol.*, *99*(3), 1366–1379. doi: 10.1152/jn.00698.2007

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network*, *12*(2), 199–213.

Chittka, L., Dyer, A. G., Bock, F., & Dornhaus, A. (2003). Bees trade off foraging speed for accuracy. *Nature*, *424*(6947), 388–388. doi:

10.1038/424388a

Chittka, L., Skorupski, P., & Raine, N. E. (2015). Speed–accuracy tradeoffs in animal decision making. *Trends Ecol. Evol.*, *24*(7), 400–407. doi: 10.1016/j.tree.2009.02.010

Churchland, A. K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., & Shadlen, M. (2011). Variance as a Signature of Neural Computations during Decision Making. *Neuron*, *69*(4), 818–831. doi: 10.1016/j.neuron.2010.12.037

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nat. Neurosci.*, *11*(6), 693–702. doi: 10.1038/nn0708-851c

Corrado, G. S., Sugrue, L. P., Seung, H. S., & Newsome, W. T. (2005). Linear-nonlinear-Poisson models of primate choice dynamics. *J. Exp. Anal. Behav.*, *84*(3), 581–617. doi: 10.1901/jeab.2005.23-05

Costa, G. M. (2015). *Neural and behavioral correlates of decision confidence* (PhD Dissertation Thesis). University of Coimbra.

Cromwell, H. C., & Schultz, W. (2003). Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J. Neurophysiol.*, *89*(5), 2823–2838. doi: 10.1152/jn.01014.2002

Cury, K. M., & Uchida, N. (2010). Robust Odor Coding via Inhalation-Coupled Transient Activity in the Mammalian Olfactory Bulb. *Neuron*, *68*(3), 570–585. doi: 10.1016/j.neuron.2010.09.040

Daniel, R., & Pollmann, S. (2012). Striatal activations signal prediction errors on confidence in the absence of external feedback. *Neuroimage*, *59*(4), 3457–3467. doi: http://dx.doi.org/10.1016/j.neuroimage.2011.11.058

DasGupta, S., Ferreira, C. H., & Miesenböck, G. (2014). FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science*, *344*(6186), 901–4. doi: 10.1126/science.1252114

Davison, M., & Baum, W. M. (2000). Choice in a variable environment: every reinforcer counts. *J. Exp. Anal. Behav.*, *74*(1), 1–24. doi: 10.1901/jeab.2000.74-1

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Curr. Opin. Neurobiol.*, *16*(2), 199–204. doi: 10.1016/j.conb.2006.03.006

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition

between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, *8*(12), 1704–1711. doi: 10.1038/nn1560

Dawkins, R. (2004). *The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution*. Houghton Mifflin.

Dayan, P., & Abbott, L. F. (2005). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press.

de C. Hamilton, A., Jones, K., & Wolpert, D. (2004). The scaling of motor noise with muscle strength and motor unit number in humans. *Exp. Brain Res.*, *157*(4), 417–430. doi: 10.1007/s00221-004-1856-7

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.*, *4*(8), 826–831. doi: 10.1038/90541

Derusso, A. L., Fan, D., Gupta, J., Shelest, O., Costa, R. M., & Yin, H. H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. *Front. Integr. Neurosci.*, *4*. doi: 10.3389/fnint.2010.00017

Dickinson, A. (1980). *Contemporary Animal Learning Theory*. Cambridge University Press.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Networks*, *19*(8), 981–1012. doi: http://dx.doi.org/10.1016/j.neunet.2006.05.042

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The Cost of Accumulating Evidence in Perceptual Decision Making. *J. Neurosci.*, *32*(11), 3612–3628. doi: 10.1523/JNEUROSCI.4010-11.2012

Duchamp-Viret, P. (1999). Odor Response Properties of Rat Olfactory Receptor Neurons. *Science (80-.).*, *284*(5423), 2171–2174. doi: 10.1126/science.284.5423.2171

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *J. Math. Psychol.*, *2*, 312–329. doi: 10.1016/0022-2496(65)90007-6

Einstein, A. (1905). On the Motion of Small Particles Suspended in a Stationary Liquid, as Required by the Molecular Kinetic Theory of Heat. *Ann. Phys.*, *322*(8), 549–560. doi: 10.1002/andp

.19053220806
Einstein, A., Podolsky, B., & Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.*, *47*(10), 777–780.
Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., & Brody, C. D. (2015). *Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat* (Vol. 4). doi: 10.7554/eLife.05457
Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi: 10.1038/415429a
Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.*, *9*(4), 292–303. doi: 10.1038/nrn2258
Fechner, G. T. (1948). *Elements of psychophysics, 1860.* East Norwalk, CT, US: Appleton-Century-Crofts. doi: 10.1037/11304-026
Feierstein, C. E., Quirk, M. C., Uchida, N., Sosulski, D. L., & Mainen, Z. F. (2006). Representation of Spatial Goals in Rat Orbitofrontal Cortex. *Neuron*, *51*(4), 495–507. doi: 10.1016/j.neuron.2006.06.032
Fellows, L. K. (2011). Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. *Ann. N. Y. Acad. Sci.*, *1239*(1), 51–58. doi: 10.1111/j.1749-6632.2011.06229.x
Feynman, R. P., Leighton, R. B., & Sands, M. L. (1963). *The Feynman Lectures on Physics* (No. v. 3). Pearson/Addison-Wesley.
FitzGerald, T. H. B., Schwartenbeck, P., & Dolan, R. J. (2014). Reward-related activity in ventral striatum is action contingent and modulated by behavioral relevance. *J. Neurosci.*, *34*(4), 1271–1279. doi: 10.1523/JNEUROSCI.4389-13.2014
Fitzpatrick, D. C., Batra, R., Stanford, T. R., & Kuwada, S. (1997). A neuronal population code for sound localization. *Nature*, *388*(6645), 871–874. doi: 10.1038/42246
Fründ, I., Wichmann, F. A., & Macke, J. H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. *J. Vis.*, *14*(7), 9. doi: 10.1167/14.7.9
Gallistel, C. R. (1993). *The Organization of Learning.* MIT Press.
Gallistel, C. R., Mark, T. A., King, A. P., & Latham, P. E. (2001). *The*

rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. (Vol. 27) (No. 4). US: American Psychological Association. doi: 10.1037/0097-7403.27.4.354

Gerwinn, S., Macke, J. H. J., & Bethge, M. (2010). Bayesian inference for generalized linear models for spiking neurons. *Front. Comput. Neurosci.*, *4*(May), 12. doi: 10.3389/fncom.2010.00012

Glimcher, P. W., & Fehr, E. (2014). Introduction: A Brief History of Neuroeconomics. In P. W. G. B. T. N. S. E. Fehr (Ed.), (pp. xvii–xxviii). San Diego: Academic Press. doi: http://dx.doi.org/10.1016/B978-0-12-416008-8.00035-8

Gold, J. I., Law, C.-T., Connolly, P., & Bennur, S. (2008). The relative influences of priors and sensory evidence on an oculomotor decision variable during perceptual learning. *J. Neurophysiol.*, *100*(5), 2653–2668. doi: 10.1152/jn.90629.2008

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*(30), 535–561. doi: 10.1146/annurev.neuro.29.051605.113038

Gottlieb, J. (2007). From thought to action: the parietal cortex as a bridge between perception, action, and cognition. *Neuron*, *53*(1), 9–16. doi: 10.1016/j.neuron.2006.12.009

Green, C. S., Benson, C., Kersten, D., & Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. *Proc. Natl. Acad. Sci.*, *107*(37), 16401–16406. doi: 10.1073/pnas.1001709107

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* John Wiley and Sons.

Gregson, R. A. M. (2014). *Time Series in Psychology.* Taylor and Francis.

Grinband, J., Hirsch, J., & Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron*, *49*(5), 757–763. doi: 10.1016/j.neuron.2006.01.032

Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, *35*(1), 4–26. doi: 10.1038/npp.2009.129

Haberly, L. B. (2001). Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry. *Chem. Senses*, *26*(5), 551–576. doi: 10.1093/chemse/26.5.551

Hanks, T., Kiani, R., & Shadlen, M. N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *Elife*, *3*, 1–17. doi: 10.7554/eLife.02260

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). *Microstimulation of macaque area LIP affects decision-making in a motion discrimination task* (Vol. 9) (No. 5). doi: 10.1038/nn1683

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, *520*(7546), 220–3. doi: 10.1038/nature14066

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *J. Neurosci.*, *28*(22), 5623–5630. doi: 10.1523/JNEUROSCI.1309-08.2008

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nat. Rev. Genet.*, *10*(8), 551–564. doi: 10.1038/nrg2593

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J.-D. (2014). The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. *Cereb. Cortex*(Dv), bhu181–. doi: 10.1093/cercor/bhu181

Hermanns, W., & Einstein, A. (1983). *Einstein and the Poet: In Search of the Cosmic Man*. Branden Press.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.*, *4*, 267–272. doi: 10.1901/jeab.1961.4-267

Herrnstein, R. J. (1970). On the law of effect. *J. Exp. Anal. Behav.*, *13*(2), 243–266. doi: 10.1901/jeab.1970.13-243

Herrnstein, R. J. (1974). Formal properties of the matching law. *J. Exp. Anal. Behav.*, *21*(1), 159–164. doi: 10.1901/jeab.1974.21-159

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. *J. Exp. Anal. Behav.*, *24*(1), 107–116. doi: 10.1901/jeab.1975.24-107

Hikosaka, K., & Watanabe, M. (2000). Delay Activity of Orbital and Lateral Prefrontal Neurons of the Monkey Varying with Different

Rewards. *Cereb. Cortex*, *10*(3), 263–271. doi: 10.1093/cercor/10.3.263

Histed, M. H., Carvalho, L. a., & Maunsell, J. H. R. (2012). Psychophysical measurement of contrast sensitivity in the behaving mouse. *J. Neurophysiol.*, *107*(3), 758–765. doi: 10.1152/jn.00609.2011

Hooper, S. L., Guschlbauer, C., von Uckermann, G., & Buschges, A. (2006). Natural neural output that produces highly variable locomotory movements. *J. Neurophysiol.*, *96*(4), 2072–2088. doi: 10.1152/jn.00366.2006

Ikemoto, S. (2007). Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res Rev*, *56*(1), 27–78. doi: 10.1016/j.brainresrev.2007.05.004

Illig, K. R. (2005). Projections from orbitofrontal cortex to anterior piriform cortex in the rat suggest a role in olfactory information processing. *J. Comp. Neurol.*, *488*(2), 224–231. doi: 10.1002/cne.20595

Ito, M., & Doya, K. (2009). Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J. Neurosci.*, *29*(31), 9861–9874. doi: 10.1523/JNEUROSCI.6157-08.2009

Ito, M., & Doya, K. (2011). Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. *Curr. Opin. Neurobiol.*, *21*(3), 368–373. doi: 10.1016/j.conb.2011.04.001

Johnson, B. A., & Leon, M. (2000). Modular representations of odorants in the glomerular layer of the rat olfactory bulb and the effects of stimulus concentration. *J. Comp. Neurol.*, *422*(4), 496–509.

Johnson, D. M., Illig, K. R., Behan, M., & Haberly, L. B. (2000). New features of connectivity in piriform cortex visualized by intracellular injection of pyramidal cells suggest that "primary" olfactory cortex functions like "association" cortex in other sensory systems. *J. Neurosci.*, *20*(18), 6974–6982.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.*, *10*(12), 1625–1633. doi: 10.1038/nn2007

Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, *63*(6), 733–745. doi: 10.1016/j.neuron.2009.09.003

Kandel, E. R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science*, *294*(5544), 1030–1038. doi: 10.1126/science.1067020

Kasamatsu, T., Polat, U., Pettet, M. W., & Norcia, A. M. (2001). Colinear facilitation promotes reliability of single-cell responses in cat striate cortex. *Exp. brain Res.*, *138*(2), 163–172.

Kasdan, L. (1994). *Wyatt Earp.*

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *J. Am. Stat. Assoc.*, *90*(430), 773–795. doi: 10.1080/01621459.1995.10476572

Kepecs, A., Uchida, N., & Mainen, Z. F. (2008). *How uncertainty boosts learning: Dynamic updating of decision strategies.* Society for Neuroscience.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227–231. doi: 10.1038/nature07200

Khan, R. M., & Sobel, N. (2004). Neural processing at the speed of smell. *Neuron*, *44*(5), 744–747.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded Integration in Parietal Cortex Underlies Decisions Even When Viewing Duration Is Dictated by the Environment. *J. Neurosci.*, *28*(12), 3017–3029. doi: 10.1523/JNEUROSCI.4761-07.2008

Kiani, R., & Shadlen, M. N. (2009). Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex. *Science (80-.).*, *324*(5928), 759–764. doi: 10.1126/science.1169405

Klasson, L., & Andersson, S. G. E. (2004). Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.*, *12*(1), 37–43. doi: 10.1016/j.tim.2003.11.006

Kording, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. doi: 10.1038/nature02169

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.*, *13*(10), 1292–1298. doi: 10.1038/nn.2635

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions. *Front. Psychol.*, *3*(June). doi: 10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl. Acad. Sci.*, *108*(33), 13852–13857. doi: 10.1073/pnas.1101328108

Lak, A., Costa, G. M., Romberg, E., Koulakov, A. a., Mainen, Z. F., & Kepecs, A. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron*, *84*(1), 190–201. doi: 10.1016/j.neuron.2014.08.039

Landau, L. D., & Lifshitz, E. M. (2013). *Statistical Physics* (No. v. 5). Elsevier Science.

Laska, M., Psychologie, M., München, L.-m.-u., & München, D. (2004). Olfactory Discrimination Ability of Human Subjects for Enantiomers with an Isopropenyl Group at the Chiral Center. *Chem. Senses*, *29*(2), 143–152. doi: 10.1093/chemse/bjh019

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *J. Exp. Anal. Behav.*, *84*(3), 555–579. doi: 10.1901/jeab.2005.110-04

Laughlin, S. B., & Sejnowski, T. J. (2003). Communication in Neuronal Networks. *Science*, *301*(5641), 1870–1874. doi: 10.1126/science.1089662

Lauwereyns, J., Watanabe, K., Coe, B., & Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature*, *418*(6896), 413–417.

Leleu, A., Demily, C., Franck, N., Durand, K., Schaal, B., & Baudouin, J.-Y. (2015). The odor context facilitates the perception of low-intensity facial expressions of emotion. *PLoS One*, *10*(9), e0138656. doi: 10.1371/journal.pone.0138656

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav. Brain Sci.*, *8*(04), 529–539.

Lindland, O. I., Sindre, G., & Solvberg, A. (1994). *Understanding quality in conceptual modeling* (Vol. 11) (No. 2). doi: 10.1109/52.268955

Link, S. W. (1992). *The Wave Theory of Difference and Similarity*. L. Erlbaum Associates.

Lottem, E., Lőrincz, M. L., & Mainen, Z. F. (2015). *Optogenetic activation of dorsal raphe serotonin neurons rapidly inhibits spontaneous but not odor-evoked activity in olfactory cortex.*

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proc. Natl. Acad. Sci. U. S. A.*, *110*(15), 6139–6144. doi: 10.1073/pnas.1217854110

Luce, R. D. (1986). *Response Times : Their Role in Inferring Elementary Mental Organization: Their Role in Inferring Elementary Mental Organization.* Oxford University Press, USA.

Lum, C. S., Zhurov, Y., Cropper, E. C., Weiss, K. R., & Brezina, V. (2005). Variability of swallowing performance in intact, freely feeding aplysia. *J. Neurophysiol.*, *94*(4), 2427–2446. doi: 10.1152/jn.00280.2005

Mackintosh, N. J. (1983). *Conditioning and associative learning.* Clarendon Press.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide.* Cambridge University Press.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In *Proc. natl. inst. sci. india* (Vol. 2, pp. 49–55).

Mannella, F., Gurney, K., & Baldassarre, G. (2013). The nucleus accumbens as a nexus between values and goals in goal-directed behavior: a review and a new hypothesis. *Front. Behav. Neurosci.*, *7*, 135. doi: 10.3389/fnbeh.2013.00135

Marcos, E., Pani, P., Brunamonti, E., Deco, G., Ferraina, S., & Verschure, P. (2013). Neural variability in premotor cortex is modulated by trial history and predicts behavioral performance. *Neuron*, *78*(2), 249–255. doi: 10.1016/j.neuron.2013.02.006

Mark, T. A., & Gallistel, C. R. (1994). *Kinetics of matching.* (Vol. 20) (No. 1). US: American Psychological Association. doi: 10.1037/0097-7403.20.1.79

Marquez, C., Rennie, S. M., Costa, D. F., & Moita, M. A. (2015). Prosocial Choice in Rats Depends on Food-Seeking Behavior Displayed by Recipients. *Curr. Biol.*, *25*(13), 1736–1745. doi: 10.1016/j.cub.2015.05.018

Martincorena, I., Seshasayee, A. S. N., & Luscombe, N. M. (2012). Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature*, *485*(7396), 95–98.

Matias, S., Lottem, E., Dugué, G. P., & Mainen, Z. F. (2015). *Sero-*

tonin and dopamine neurons signal distinct prediction errors during reversal learning.

McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y., & Schoenbaum, G. (2011). Ventral Striatum and Orbitofrontal Cortex Are Both Required for Model-Based, But Not Model-Free, Reinforcement Learning. *J. Neurosci.*, *31*(7), 2700–2705. doi: 10.1523/JNEUROSCI.5499-10.2011

Michel, M. M., & Jacobs, R. A. (2008). Learning optimal integration of arbitrary features in a perceptual discrimination task. *J. Vis.*, *8*(2), 3.1–16. doi: 10.1167/8.2.3

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The Drift Diffusion Model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgm. Decis. Mak.*, *5*(6), 437–449.

Miura, K., Mainen, Z. F., & Uchida, N. (2012). Odor Representations in Olfactory Cortex: Distributed Rate Coding and Decorrelated Population Activity. *Neuron*, *74*(6), 1087–1098. doi: 10.1016/j.neuron.2012.04.021

Morrison, S. E., & Salzman, C. D. (2011). Representations of appetitive and aversive information in the primate orbitofrontal cortex. *Ann. N. Y. Acad. Sci.*, *1239*(1), 59–70. doi: 10.1111/j.1749-6632.2011.06255.x

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the Brain: A Diffusion Model Analysis of Prior Probability and Potential Payoff. *J. Neurosci.*, *32*(7), 2335–2343. doi: 10.1523/JNEUROSCI.4156-11.2012

Murakami, M., Kashiwadani, H., Kirino, Y., & Mori, K. (2005). State-dependent sensory gating in olfactory cortex. *Neuron*, *46*(2), 285–296. doi: 10.1016/j.neuron.2005.02.025

Murakami, M., Vicente, M. I., Costa, G. M., & Mainen, Z. F. (2014). Neural antecedents of self-initiated actions in secondary motor cortex. *Nat Neurosci*, *17*(11), 1574–1582. doi: 10.1038/nn.3826

Murray, E. A., O'Doherty, J. P., & Schoenbaum, G. (2007). What we know and do not know about the functions of the orbitofrontal cortex after 20 years of cross-species studies. *J. Neurosci.*, *27*(31), 8166–9. doi: 10.1523/JNEUROSCI.1556-07.2007

Nomoto, K., & Lima, S. Q. (2015). Enhanced male-evoked responses in the ventromedial hypothalamus of sexually receptive female mice. *Curr. Biol.*, *25*(5), 589–594. doi: 10.1016/j.cub.2014.12.048

Orsborn, A. L., & Carmena, J. M. (2013). Creating new functional circuits for action via brain-machine interfaces. *Front. Comput. Neurosci.*, *7*, 157. doi: 10.3389/fncom.2013.00157

Osborne, L. C., Lisberger, S. G., & Bialek, W. (2005). A sensory source for motor variation. *Nature*, *437*(7057), 412–416. doi: 10.1038/nature03961

Padoa-Schioppa, C., & Assad, J. a. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226. doi: 10.1038/nature04676

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.*, *11*(1), 95–102. doi: 10.1038/nn2020

Padoa-Schioppa, C., & Cai, X. (2011). Orbitofrontal Cortex and the Computation of Subjective Value: Consolidated Concepts and New Perspectives. *Ann. N. Y. Acad. Sci.*, *1239*, 130–137. doi: 10.1111/j.1749-6632.2011.06262.x

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.*, *5*(5), 1–1. doi: 10.1167/5.5.1

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, *439*(7078), 865–870. doi: 10.1038/nature04490

Patterson, M. A., Lagier, S., & Carleton, A. (2013). Odor representations in the olfactory bulb evolve after the first breath and persist as an odor afterimage. *Proc Natl Acad Sci U S A*, *110*(35), E3340-9. doi: 10.1073/pnas.1303873110

Pierce, J. D., Zeng, X.-N., Aronov, E. V., Preti, G., & Wysocki, C. J. (1995). Cross-adaptation of Sweaty-smelling 3-methyl-2- hexenoic Acid by a Structurally-similar, Pleasant-smelling Odorant. *Chem. Senses*, *20*(4), 401–411. doi: 10.1093/chemse/20.4.401

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008). Spatio-temporal correlations

and visual signalling in a complete neuronal population. *Nature*, *454*(7207), 995–999. doi: 10.1038/nature07140

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.*, *85*(2), 59–108. doi: 10.1037/0033-295X.85.2.59

Ratcliff, R., & McKoon, G. (2008). *The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks* (Vol. 20) (No. 4). doi: 10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychol. Sci.*, *9*(5), 347–356. doi: 10.1111/1467-9280.00067

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.*, *111*, 333–367. doi: 10.1037/0033-295X.111.2.333

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). *Connectionist and diffusion models of reaction time.* (Vol. 106) (No. 2). US: American Psychological Association. doi: 10.1037/0033-295X.106.2.261

Reisert, J., & Matthews, H. R. (2000). Adaptation-induced changes in sensitivity in frog olfactory receptor cells. *Chem. Senses*, *25*(4), 483–486.

Reisert, J., & Matthews, H. R. (2001). Response properties of isolated mouse olfactory receptor cells. *J. Physiol.*, *530*(Pt 1), 113–122.

Rennaker, R. L., Chen, C.-F. F., Ruyle, A. M., Sloan, A. M., & Wilson, D. A. (2007). Spatial and Temporal Distribution of Odorant-Evoked Activity in the Piriform Cortex. *J. Neurosci.*, *27*(7), 1534–1542. doi: 10.1523/JNEUROSCI.4072-06.2007

Rescorla, R. A. (1988). Pavlovian Conditioning: It's Not What You Think It Is. *Acta Biol Exp (Warsz).*, *43*(3), 151–160.

Rescorla, R. a., & Wagner, a. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Class. cond. ii curr. res. theory* (Vol. 21, pp. 64–99). New York: Appleton-Century-Crofts. doi: 10.1101/gr.110528.110

Rinberg, D., Koulakov, A., & Gelperin, A. (2006a). Sparse odor coding in awake behaving mice. *J. Neurosci.*, *26*(34), 8857–8865. doi: 10.1523/JNEUROSCI.0884-06.2006

Rinberg, D., Koulakov, A., & Gelperin, A. (2006b). Speed-Accuracy

Tradeoff in Olfaction. *Neuron*, *51*(3), 351–358. doi: 10.1016/j.neuron.2006.07.013

Roesch, M., Taylor, A., & Schoenbaum, G. (2006). Encoding of Time-Discounted Rewards in Orbitofrontal Cortex Is Independent of Value Representation. *Neuron*, *51*(4), 509–520. doi: 10.1016/j.neuron.2006.06.027

Roesch, M. R., & Olson, C. R. (2004). Neuronal Activity Related to Reward Value and Motivation in Primate Frontal Cortex. *Science (80-.).*, *304*(5668), 307–310. doi: 10.1126/science.1093223

Roesch, M. R., Singh, T., Brown, P. L., Mullins, S. E., & Schoenbaum, G. (2009). Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *J. Neurosci.*, *29*(42), 13365–13376. doi: 10.1523/JNEUROSCI.2572-09.2009

Roesch, M. R., Stalnaker, T. A., & Schoenbaum, G. (2006). Associative Encoding in Anterior Piriform Cortex versus Orbitofrontal Cortex during Odor Discrimination and Reversal Learning. *Cereb. Cortex*, *17*(3), 643–652. doi: 10.1093/cercor/bhk009

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.*, *22*(21), 9475–9489. doi: 10.1016/S0377-2217(02)00363-6

Rolls, E., & Deco, G. (2010). The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function. *Anim. Behav.*, *80*(1), 171. doi: 10.1016/j.anbehav.2010.04.019

Rushworth, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, *70*(6), 1054–1069. doi: 10.1016/j.neuron.2011.05.014

Sajda, P., Philiastides, M. G., Heekeren, H., & Ratcliff, R. (2011). *Linking Neuronal Variability to Perceptual Decision Making via Neuroimaging.* doi: 10.1093/acprof:oso/9780195393798.001.0001

Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S., & Quinn, J. T. J. (1979). Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychol. Rev.*, *47*(5), 415–451.

Schoenbaum, G., Takahashi, Y., Liu, T.-L., & McDannald, M. A. (2011).

Does the orbitofrontal cortex signal value? *Ann. N. Y. Acad. Sci.*, *1239*, 87–99. doi: 10.1111/j.1749-6632.2011.06210.x

Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiol Rev*, *95*(3), 853–951. doi: 10.1152/physrev.00023.2014

Schultz, W., Tremblay, L., & Hollerman, J. R. (2000). Reward Processing in Primate Orbitofrontal Cortex and Basal Ganglia. *Cereb. Cortex*, *10*(3), 272–283.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.*, *6*(2), 461–464. doi: 10.1214/aos/1176344136

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.*, *16*(4), 1486–1510.

Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*(3), 791–806. doi: 10.1016/j.neuron.2013.10.047

Shanks, D. R. (1995). *The Psychology of Associative Learning*. Cambridge University Press.

Shusterman, R., Smear, M. C., Koulakov, A. A., & Rinberg, D. (2011). Precise olfactory responses tile the sniff cycle. *Nat Neurosci*, *14*(8), 1039–1044. doi: 10.1038/nn.2877

Simmons, J. M., Ravel, S., Shidara, M., & Richmond, B. J. (2007). A comparison of reward-contingent neuronal activity in monkey orbitofrontal cortex and ventral striatum: guiding actions toward rewards. *Ann. N. Y. Acad. Sci.*, *1121*, 376–94. doi: 10.1196/annals.1401.028

Simoncelli, E. P., Paninski, L., Pillow, J., & Schwartz, O. (2004). Characterization of Neural Responses with Stochastic Stimuli. *Cogn. Neurosci. Iii, Third Ed.*, 327–338. doi: 10.1161/CIRCULATIONAHA.108.827121.A

Smith, P. L. (1994). Fechner's Legacy and Challenge. Review of The Wave Theory of Difference and Similarity, by Stephen W. Link. *J. Math. Psychol.*, *38*(3), 407–420. doi: http://dx.doi.org/10.1006/jmps.1994.1029

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice

discrimination. *J. Math. Psychol.*, *32*(2), 135–168. doi: 10.1016/0022-2496(88)90043-0

Spors, H., & Grinvald, A. (2002). Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. *Neuron*, *34*(2), 301–315.

Stages, L. (2002). Letters To Nature. *Ecology*, *418*(JULY), 413–417. doi: 10.1038/nature00844.1.

Stettler, D. D., & Axel, R. (2009). Representations of Odor in the Piriform Cortex. *Neuron*, *63*(6), 854–864. doi: 10.1016/j.neuron.2009.09.005

Stevens, S. S. (1957). *On the psychophysical law.* (Vol. 64) (No. 3). US: American Psychological Association. doi: 10.1037/h0046162

Stevens, S. S. (1975). *Psychophysics.* Transaction Publishers.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.*, *9*(4), 578–585. doi: 10.1038/nn1669

Stüttgen, M. C. (2011). Mapping spikes to sensations. *Front. Neurosci.*, *5*(November), 1–17. doi: 10.3389/fnins.2011.00125

Summerfield, C., Behrens, T. E., & Koechlin, E. (2011). Perceptual classification in a rapidly changing environment. *Neuron*, *71*(4), 725–736. doi: 10.1016/j.neuron.2011.06.022

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.*, *15*(11), 745–756. doi: 10.1038/nrn3838

Summerfield, C., & Tsetsos, K. (2012). Building Bridges between Perceptual and Economic Decision-Making: Neural and Computational Mechanisms. *Front. Neurosci.*, *6*, 70. doi: 10.3389/fnins.2012.00070

Summerfield, C., & Tsetsos, K. (2015). Do humans make good decisions? *Trends Cogn. Sci.*, *19*(1), 27–34. doi: 10.1016/j.tics.2014.11.005

Sutton, R. S., & Barto, A. G. (1990). Time-Derivative Models of Pavlovian Reinforcement. *Learn. Comput. Neurosci. Found. Adapt. Networks*(Mowrer 1960), 497–537. doi: 10.1111/j.1748-1716.1960.tb01900.x

Sutton, R. S., & Barto, A. G. (1998). *Introduction to Reinforcement Learning* (1st ed.). Cambridge, MA, USA: MIT Press.

Takahashi, Y. K., Chang, C. Y., Lucantonio, F., Haney, R. Z., Berg, B. A., Yau, H.-J., ... Schoenbaum, G. (2013). Neural estimates

of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron*, *80*(2), 10.1016/j.neuron.2013.08.008. doi: 10.1016/j.neuron.2013.08.008

Taniguchi, M., Kashiwayanagi, M., & Kurihara, K. (1992). Quantitative analysis on odor intensity and quality of optical isomers in turtle olfactory system. *Am. J. Physiol. - Regul. Integr. Comp. Physiol.*, *262*(1), R99–R104.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic Modeling of Elementary Psychological Processes*. Cambridge University Press.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects. *J. Neurophysiol.*, *93*(2), 1074–1089.

Uchida, N., Kepecs, A., & Mainen, Z. F. (2006). Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. *Nat. Rev. Neurosci.*, *7*(6), 485–491. doi: 10.1038/nrn1933

Uchida, N., & Mainen, Z. F. (2003). Speed and accuracy of olfactory discrimination in the rat. *Nat. Neurosci.*, *6*(11), 1224–1229. doi: 10.1038/nn1142

Uchida, N., Poo, C., & Haddad, R. (2014). Coding and Transformations in the Olfactory System. *Annu. Rev. Neurosci.*, *37*(1), 363–385. doi: 10.1146/annurev-neuro-071013-013941

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychol. Rev.*, *108*(3), 550–592. doi: 10.1037//0033-295X.108.3.550

van der Meer, M. A. A., & Redish, A. D. (2009). Covert Expectation-of-Reward in Rat Ventral Striatum at Decision Points. *Front. Integr. Neurosci.*, *3*, 1. doi: 10.3389/neuro.07.001.2009

van der Meer, M. A. A., & Redish, A. D. (2011). Ventral striatum: a critical look at models of learning and evaluation. *Curr. Opin. Neurobiol.*, *21*(3), 387–392. doi: 10.1016/j.conb.2011.02.011

Vicente, M. I. (2015). *Uncertainty in olfactory decision-making* (PhD Dissertation Thesis). Universidade Nova de Lisboa.

Vickers, D. (1970). Evidence for an Accumulator Model of Psychophysical Discrimination. *Ergonomics*, *13*(1), 37–58. doi: 10.1080/00140137008931117

Vickers, D., Carterette, E. C., & Friedman, M. P. (2014). *Decision Processes in Visual Perception.* Elsevier Science.

Von Neumann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Autom. Stud.*, *34*, 43–98.

Waiblinger, C., Brugger, D., Whitmire, C. J., Stanley, G. B., & Schwarz, C. (2015). *Support for the slip hypothesis from whisker-related tactile perception of rats in a noisy environment* (Vol. 9).

Wald, A., & Wolfowitz, J. (1949). Bayes Solutions of Sequential Decision Problems. *Proc. Natl. Acad. Sci. U. S. A.*, *35*(2), 99–102.

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.*, *30*, 31–56. doi: 10.1146/annurev.neuro.30.051606.094334

Wallis, J. D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat Neurosci*, *15*(1), 13–19.

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.*, *18*(7), 2069–2081. doi: 10.1046/j.1460-9568.2003.02922.x

Wesson, D. W., Verhagen, J. V., & Wachowiak, M. (2009). Why sniff fast? the relationship between sniff frequency, odor discrimination, and receptor neuron activation in the rat. *J Neurophysiol*, *101*(2), 1089–1102. doi: 10.1152/jn.90981.2008

Wesson, D. W., & Wilson, D. A. (2011). Sniffing out the contributions of the olfactory tubercle to the sense of smell: hedonics, sensory integration, and more? *Neurosci Biobehav Rev*, *35*(3), 655–668. doi: 10.1016/j.neubiorev.2010.08.004

Wichmann, F., & Hill, N. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.*, *63*(8), 1293–1313. doi: 10.3758/BF03194544

Widrow, B., & Hoff, M. (1960). *Adaptive switching circuits.* (No. 4).

Wilson, D. A. (2000). Comparison of odor receptive field plasticity in the rat olfactory bulb and anterior piriform cortex. *J. Neurophysiol.*, *84*(6), 3036–3042.

Wilson, D. A. (2003). Rapid, Experience-Induced Enhancement in Odorant Discrimination by Anterior Piriform Cortex Neurons. *J. Neurophysiol.*, *90*(1), 65–72. doi: 10.1152/jn.00133.2003

Wit, E., van den Heuvel, E., & Romeijn, J.-W. (2012). 'All models are wrong...': an introduction to model uncertainty. *Stat. Neerl.*, *66*(3), 217–236. doi: 10.1111/j.1467-9574.2012.00530.x

Wojcik, P. T., & Sirotin, Y. B. (2015). Single Scale for Odor Intensity in Rat Olfaction. *Curr. Biol.*, *24*(5), 568–573. doi: 10.1016/j.cub.2014.01.059

Wright, B. E. (2000). A Biochemical Mechanism for Nonrandom Mutations and Evolution. *J. Bacteriol.*, *182*(11), 2993–3001. doi: 10.1128/JB.182.11.2993-3001.2000.Updated

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075–1080. doi: 10.1038/nature05852

Yu, A. J., & Cohen, J. D. (2008). Sequential effects: Superstition or rational behavior? *Adv. Neural Inf. Process. Syst.*, *21*, 1873–1880.

Zariwala, H. A., Kepecs, A., Uchida, N., Hirokawa, J., & Mainen, Z. F. (2013). The Limits of Deliberation in a Perceptual Decision Task. *Neuron*, *78*(2), 339–351. doi: 10.1016/j.neuron.2013.02.010

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*(6485), 140–143. doi: 10.1038/370140a0

Appendices

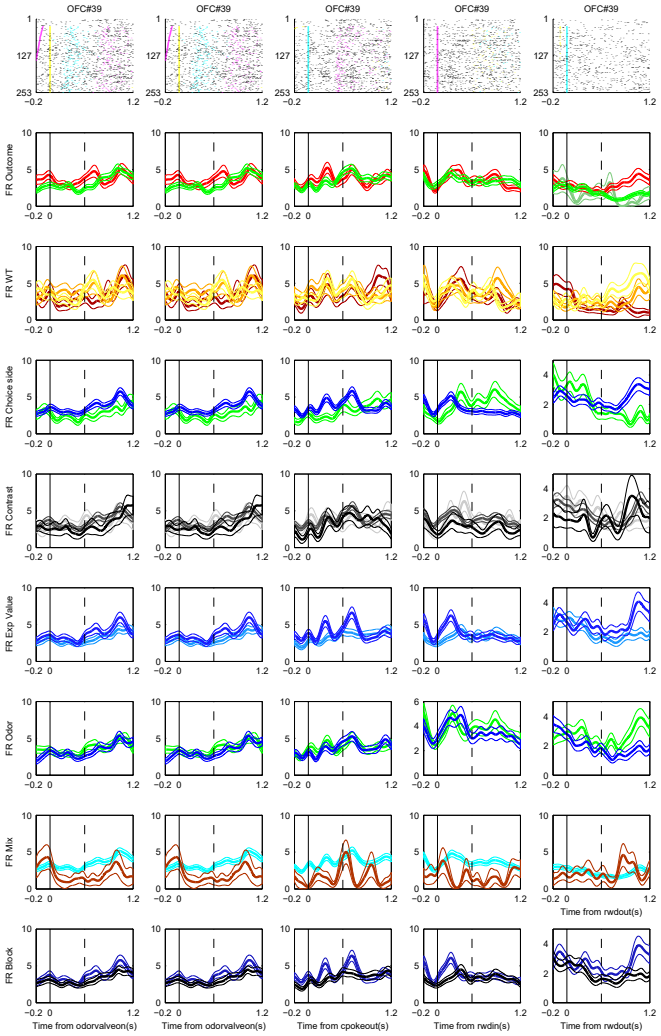

Figure 2. PETHs for OFC-39.

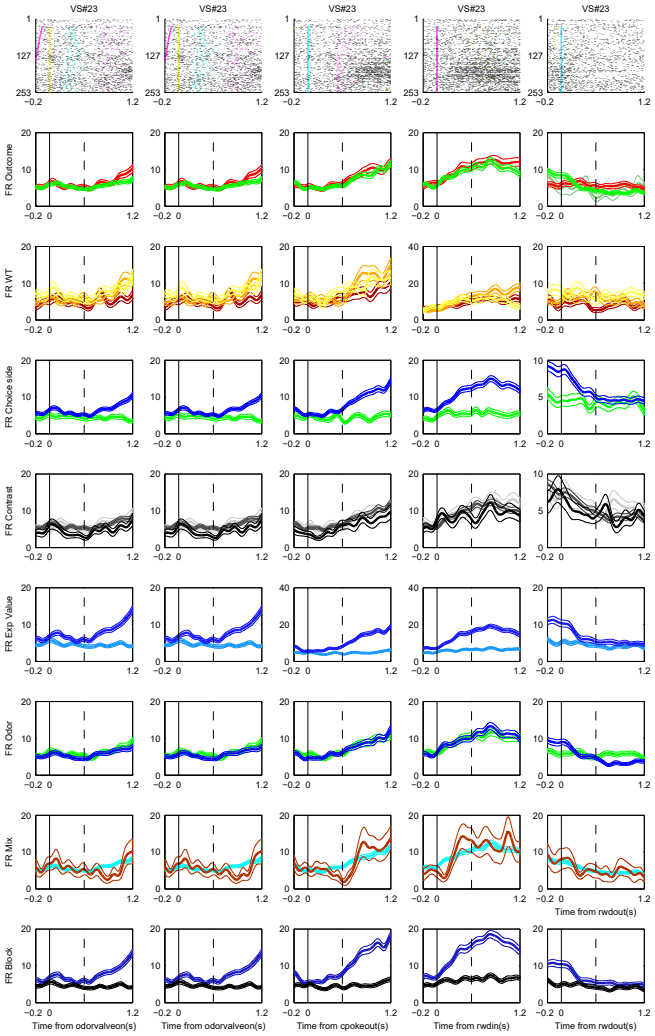

Figure 3. PETHs for VS-23.

"Conglaturation! You have completed a great game! And prooved the justice of our culture."
– Ending screen, *Ghostbusters*

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt