

Gil Miguel Pereira da Costa

Neural and behavioral correlates of decision confidence

PhD in Biosciences, branch of specialization in Neurosciences.

This work was developed under the academic supervision of Zachary F. Mainen, PhD, Champalimaud Neuroscience Program, Lisbon, and Carlos B. Duarte, PhD, Department of Life Sciences, Faculty of Sciences and Technology, Universidade de Coimbra, Coimbra, Portugal, and presented to the Faculty of Sciences and Technology, University of Coimbra.

September 2014



UNIVERSIDADE DE COIMBRA

a minha mãe, pela força

This work was developed in the context of the PhD in Biosciences, branch of specialization in Neurosciences, at the Center for Neuroscience and Cell Biology of Coimbra, University of Coimbra, Coimbra, Portugal, under the scientific supervision of Professor Carlos Duarte. The project entitled “Neural and Behavioral Correlates of Decision Confidence” was carried out at the Instituto Gulbenkian de Ciência, Oeiras, Portugal; at the Department of Neurosciences of the Cold Spring Harbor Laboratory, New York, USA and at the Champalimaud Neuroscience Program, Champalimaud Centre for the Unknown, Lisbon Portugal, under the scientific supervision of Professor Zachary Mainen. This work was supported by the fellowship SFRH / BD / 32947 / 2006 from the Foundation for Science and Technology, Portugal.

ACKNOWLEDGEMENTS

To Maria and Masa, for the support.

To Zach, for the wings to fly.

To Patrícia for complicity.

To André and Rita, for friendship.

To Nicco, Eran, Eric, Bass, Cindy, Rita, Sara, Madalena, Paul and the rest of the gang for, making it a special group.

To Zach, Marta, Rui, Susana and Joe, for making CNP happen. And to the rest of CNP, for the diversity, vibe and power to achieve.

To Gonzalo, Florin, Armin, Adam, Barry, Josh, Asheesh, Margarida, Mariana, Hassana and Peter, for friendship in distant places.

To Daniel, for friendship and example.

To Joana, Vera, Sara, Marina and Ana, for the early days.

To Carlos Duarte and João Ramalho, for the help with Coimbra.

To my family, and friends, for the roots.

To Olga, my mother, for the gift of strength.

To Fernando, my father, for the gift of culture.

To my grandparents, João, Lurdes, Ester, for love and a past.

To aunt Guida, for showing will and interest.

To João, for the gift of unconditional brotherhood.

To Joana, Ulisses and Alix for the gift of timeless friendship.

To Teresa, Vicente, Rita and Alice, for companionship.

To Carolina, for love and sorrow.

Thank you.

RESUMO

Os juízos de confiança são exemplos centrais da metacognição – conhecimento acerca dos processos cognitivos de cada um. De acordo com a visão metacognitiva, os relatos de confiança são gerados por um processo de monitorização de segunda-ordem, baseado na qualidade das representações internas sobre crenças próprias. Apesar dos correlatos neurais de confiança nas decisões terem sido recentemente identificados em humanos e outros animais, não é ainda muito claro se existem regiões do cérebro especialmente importantes na monitorização da confiança.

De forma a explorar este assunto criámos uma tarefa comportamental de aposta temporal pós-decisão onde ratos expressam confiança relacionada com a escolha, através da quantidade de tempo que estão dispostos a esperar por recompensas. Nesta tarefa os ratos tiveram de escolher entre duas opções - deslocarem-se para uma zona à sua esquerda ou à sua direita. A evidência para efectuar uma decisão correcta era dada por um estímulo odorífero. Depois de uma decisão ter sido feita os animais esperavam por recompensa. Estavam dispostos a esperar mais tempo por uma recompensa quando previam que a decisão fosse acertada. Este tempo era também dependente da dificuldade da decisão – tempos de espera maiores foram observados em decisões fáceis e correctas.

Posteriormente demonstrámos que um modelo normativo conseguia prever quantitativamente os tempos de espera, baseados na computação da confiança nas decisões, estabelecendo um sistema computacional para o estudo de relatos de confiança, que descarta descrições semânticas.

De seguida inactivámos farmacologicamente o córtex orbitofrontal (OFC) de ratos enquanto estes efectuavam a tarefa comportamental de “tempo de espera”. Esta manipulação perturbou os relatos de confiança baseados no tempo de espera que se tornou menos dependente do facto da decisão ser ou não correcta, e também da sua dificuldade. No entanto

a manipulação não alterou a performance dos animais. Estes resultados estabelecem um locus anatómico para relatos cognitivos, julgamentos de confiança, distinto dos processos necessários para formar decisões perceptuais.

Decisões difíceis podem ocorrer porque os estímulos são difíceis de percepcionar ou porque as regras que definem o que se deve fazer em resposta aos estímulos apresentados são incertas para o decisor. Nós gostaríamos de perceber melhor como é que esta segunda forma de incerteza é representada no cérebro e pode ser avaliada e usada para comportamentos adaptativos. Correlatos neurais de confiança em decisões perceptuais foram já descobertos no OFC de ratos. O OFC e o estriado ventral (VS) são duas regiões cerebrais implicadas na supervisão comportamental e avaliação de resultados.

Para melhor perceber o papel do OFC e do VS na computação da confiança nas decisões e na adaptação comportamental sintonizada por sinais de confiança nós registámos a actividade de neurónios individuais situados nestas duas regiões, enquanto ratos efectuavam a tarefa comportamental de tempo de espera. Encontrámos populações de neurónios, em ambas as regiões, cuja actividade estava correlacionada com a confiança nas decisões e com o tempo de espera, logo após os ratos terem feito a sua escolha. Estes resultados exploraram mais além a função do OFC em decisões baseadas no nível de confiança, e adicionaram os gânglios de base ao circuito envolvido na computação de confiança nas decisões.

As decisões perceptuais categóricas acontecem baseadas tanto na informação sensory como em factores relacionados com o reforço, tais como os valores da recompensa. Quando uma decisão é incerta os animais enviesam as suas escolhas a favor da opção mais reforçadora. Se por um lado é sabido que as escolhas são enviesadas pelo valor não é claro como é que a magnitude da recompensa afecta decisões baseadas na confiança.

Para melhor perceber o efeito da recompensa em julgamentos de confiança após a decisão treinámos ratos para efectuar uma versão modificada da tarefa do tempo de espera, com uma manipulação de recompensa. Em conjuntos de ensaios era oferecido aos animais uma maior quantidade de recompensa por cada decisão correcta, para uma das localizações. Observámos que a manipulação da recompensa enviesou a escolha dos animais na direcção da opção mais recompensadora. Esta manipulação também afectou a sua performance. Decisões para o lado oposto à opção mais recompensadora foram mais correctas. Não obstante, esta manipulação da quantidade de recompensa não alterou os relatos de confiança baseados no tempo de espera. Os ratos esperavam a mesma quantidade de tempo enquanto esperavam pela recompensa maior ou pela recompensa menor.

Posteriormente conseguimos explicar os resultados comportamentais de performance, escolha e tempo de espera, formulando um modelo baseado em princípios de teoria de detecção de sinal. Neste modelo dois factores opostos – uma estimativa de confiança enviesada pela recompensa, e outra estimativa de confiança não enviesada – interagem para dar lugar a um relato de confiança. Esta estrutura teórica explora também a hipótese de que existe uma distinção entre decisões perceptuais e estimativas de confiança.

Os resultados publicados nesta dissertação podem ser aplicados na construção de teorias normativas de tomadas de decisão por humanos, e constroem o circuito neuronal envolvido na produção de julgamentos de confiança. Estes resultados trazem implicações na forma como percebemos como é que as pessoas avaliam as suas decisões e adaptam e optimizam o seu comportamento face a um mundo incerto, complexo e pleno em mudança.

SUMMARY

Confidence judgments are a central example of metacognition—knowledge about one's own cognitive processes. According to this metacognitive view, confidence reports are generated by a second-order monitoring process based on the quality of internal representations about beliefs. Although neural correlates of decision confidence have been recently identified in humans and other animals, it is not well understood whether there are brain areas specifically important for confidence monitoring. To address this issue we designed a post-decision temporal wagering task in which rats expressed choice confidence by the amount of time they were willing to wait for rewards.

In this task rats had to choose amongst two options, going to a choice location either on the left or right. Evidence for correctness of choice was given by a odor stimulus. After a decision was made rats would have to wait for reward. They were willing to wait longer for a reward when expecting it to be correct. And this was difficulty dependent - longer waiting times were observed after correct easy choices. Furthermore we have shown that a normative model can quantitatively account for waiting times based on the computation of decision confidence, establishing a computational framework for studying decision reports, which puts aside semantic descriptions.

Next we have pharmacologically inactivated the orbitofrontal cortex (OFC) of rats performing the waiting time task. This inactivation disrupted waiting-based confidence reports - waiting time was less dependent of correctness and difficulty of decision. But this manipulation did not affect decision accuracy. These results establish an anatomical locus for a metacognitive report, confidence judgment, distinct from the processes required for perceptual decisions.

Difficult decisions can occur because stimuli are hard to perceive or because the rules of what should be done given a certain stimulus are uncertain to the decision maker. We would like to understand how this

second form of uncertainty is represented by the brain and may be assessed and used for adaptive behavior. Neural correlates of perceptual decision confidence have been previously found in the OFC of rats. OFC and ventral striatum (VS) are two brain regions implicated in behavioral supervision and outcome evaluation.

To better understand the role of OFC and VS in the computation of decision confidence and behavior adaptation tuned by confidence signals we have recorded single unit activity from these two regions from rats performing the waiting time task. We have found populations of cells in both OFC and VS whose activity was correlated with decision confidence and waiting time, soon after decision was made. These results have further explored the functions of OFC in confidence based guided decisions and added the basal ganglia to the circuitry involved in decision confidence computations.

Perceptual categorical decisions take place based both in stimulus information and reinforcement-related factors, such as the value of outcomes. When a decision is uncertain animals bias their choices in favor of the most rewarding option. While choices can be biased by value, it is not clear if reward magnitude affects decisions based on confidence.

To better understand the effect of reward in post-decision confidence judgments we have trained rats to perform a modified version of the waiting task with a block-wise reward manipulation. In blocks of trials animals were offered a higher reward amount for one of the options. We have observed that the reward manipulation biased animals choices towards the most rewarding option. It also affected their performance. Decisions away from the highest rewarding option were more accurate. Nevertheless this reward manipulation did not alter the waiting time confidence report. Rats were waiting the same amount of time when expecting a smaller or a larger amount of reward.

By devising a signal-detection-theory based model which assumed a dual-route processing of confidence we were able to explain, our

accuracy, choice and waiting time behavioural results. In this model two opposite factors - a confidence estimate which is biased by reward, and a unbiased confidence estimate - interact to give rise to a confidence report. This theoretical framework further explores a distinction between perceptual decisions and confidence estimates.

Taken together the results portrayed in this dissertation can be applied to build up normative theories of human decision-making and establish the neural circuitry involved in producing confidence judgments. The results have implications in our understanding of how do people evaluate their decisions and can further adapt and optimize behavior in the face of an uncertain, complex and ever-changing environment.

TABLE OF CONTENTS

Acknowledgments	iv
Resumo	v
Summary	viii
<hr/>	
THESIS INTRODUCTION	1
Decisions: lenses to better understanding the brain	2
Neural mechanisms of decision making	8
Aims	20
<hr/>	
CHAPTER 1 - ORBITOFRONTAL CORTEX IS REQUIRED FOR OPTIMAL WAITING BASED ON DECISION CONFIDENCE	21
Summary	22
Introduction	22
Results	24
Discussion	42
Experimental procedures I	47
<hr/>	
CHAPTER 2 -NEURAL CORRELATES OF CONFIDENCE AND CONFIDENCE REPORT IN ORBITOFRONTAL CORTEX AND VENTRAL STRIATUM	56
Summary	57
Introduction	57
Results	60
Discussion	73
<hr/>	
CHAPTER 3 -INTERPLAY BETWEEN CONFIDENCE AND VALUE TO GENERATE A CONFIDENCE REPORT	78
Summary	79
Introduction	79
Results	82
Discussion	92
Experimental procedures II & III	98
<hr/>	
THESIS DISCUSSION	109
<hr/>	
REFERENCES	118

INTRODUCTION

“Brains allow us to perceive the world, respond to it, move through it, and act on it.” (Greenspan, 2007).

“Marr said that “an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (the hardware) in which is embodied”.[(Marr, 1982)] I want to suggest that at a global level we can characterize the function of the nervous system as decision making.”(Glimcher, 2003)

“When we perceive the physical world, make a decision, and take an action a critical issue that our brains must deal with is uncertainty.” (Doya, Ishii, Pouget, & Rao, 2007).

Decisions: lenses to better understanding the brain

It is humbling to notice that “perception” of a very simple stimulus, and a stereotypical response to it, could be achieved by a simple biochemical and molecular machinery, like the one from *Euglena*, a simple unicellular aquatic protozoan (Swanson, 2003). Because in this organism sensors and effectors are located in one single cell, the types of “behaviors” produced are very primitive. But given that the world is much more complex than a quiet, constrained water pond, it is fortunate for us humans that along natural history the complexity of nervous system organization has evolved dramatically. Otherwise we would have to struggle with, and as (un) creatively as protozoans for the same ecological niche.

Our brain, with all its different layers of neurons that build up into complex systems, allows us to cope with a more dynamic and uncertain world. Our behavioural repertoire makes a big difference. We don’t passively “deal” with the world, we are able to act upon and adapt to it by the virtue of our decisions.

In the hope to understand the nervous system we should look to understand its function. The study of decision-making provides insight into the realm of brain functions that comprise cognition, linking perception - the processing of sensory stimuli - to the output of motor actions, which unfolds into behavior. Decision-making conjugates the fields of cognitive neurosciences, systems neurosciences and computational neurosciences into better defining the neuronal circuits and computations that ultimately bring “thinking” to the cellular scale.

A first glimpse into uncertainty

There are four basic computations in the core of most decision processes, which have been the focus for understanding the neural underpinnings of decision-making: accumulation of evidence, formation of a categorical choice, reward-based adaptation and stochasticity inherent in choice

behavior (Wang, 2008). We are mostly interested in the latter one and in questions that follow from it.

Why do we make incorrect decisions? This comes from the fact that the decision process is noisy (Knill and Pouget, 2004), so uncertainty is to blame. Noise in the nervous system might have multiple sources (Faisal et al., 2008) but we consider that uncertainty in categorical decision making mainly comes from: 1) uncertainty arising from the outside world (not being sure about the information we obtain from our senses), and/or 2) internal uncertainty (not being sure on what to do with the information we capture).

Given that uncertainty in decisions might be ubiquitous it is feasible to assume that the nervous system has computational strategies and appropriate circuitry to calculate the amount of uncertainty in a decision, and to take advantage of these computations to optimally drive and adapt behaviors (Drugowitsch et al., 2012; Kording, 2007). For instance, when one is uncertain, learning rates should be boosted and attention enhanced (Dayan et al., 2000). One other function of uncertainty, which we will focus on, is that it could be used to compute confidence judgments, used to evaluate outcomes of decisions before they are known, and drive optimal post-decision strategies (Kepcs and Mainen, 2012).

We shall focus on perceptual decisions, that involve categorical judgments on the basis of sensory stimuli. These are deliberative processes that combine the available sensory evidence with information related to the known alternatives, past experiences, and decision goals to drive decisions (Ding and Gold, 2013).

From the senses towards reward

What is the purpose of choosing? What is the goal of a decision? For the argument of simplicity let us assume that decisions are made in order to obtain desired outcomes, and avoid undesired ones. These contribute to establish a decision rule – *given what I have and what I want, what should I do?*

In the end-side of a perceptual-decision lies a goal, but it all starts with a stimulus. To understand the neural circuits and computations giving rise to perceptual decisions, one must describe how these decisions depend on sensory input. Psychophysics is a field developed to describe just that. By varying characteristics of presented stimuli, such as contrast or intensities, it is possible to manipulate the amount of sensory evidence available for each individual decision, and relate perception with choice behavior. To prevent interpretational confounds due to choice bias from stimulus attributes unrelated to sensory features (like different motivational values) it is important to correctly design a behavioral task (for a simple review, Carandini & Churchland, 2013). The two-alternative-forced-categorization task design can be good to prevent these confounds.

In the context of most decisions a desired outcome is some “reward”, which has a given value. Animals make decisions for food (eg. Smith et al., 1995), drink (eg. Padoa-Schioppa & Cai, 2011), sexual content (eg. Deaner, Khera, & Platt, 2005) or social interactions (eg. Zinck & Lima, 2013), because these have some positive subjective value to the animal, and avoid punishment (eg. Paton, Belova, Morrison, & Salzman, 2006) because it has a negative value. A basic premise is that decisions are influenced by the net value that they might bring to an organism that aims to satisfy its needs. By this premise, an animal will choose to perform certain actions, in detriment of others, so to maximize rate of obtained reward (Herrnstein, 1961).

Value is a concept that has been key in the study of behavioural decision-making by multiple fields. For instance, in machine learning theory (Barto and Sutton, 1998) maximizing ‘state’ values and ‘action’ values leads to learning and selection of actions. In neuroeconomics (Glimcher and Rustichini, 2004) assignment of economic value allows qualitatively different goods to be compared in a ‘universal currency’. In animal learning theory (Balleine and Dickinson, 1998) the incentive value of outcomes is responsible for behavioural motivation.

In order to study the neural computations of decisions, animals are trained to perform behavioural tasks where the input variables are well controlled in the laboratory environment, and the behavior of the animal is characterized to a great extent. Traditionally, behavior is discretized into trials and neuronal activity can be aligned to specific events within individual trials. The variables of interest can vary. For instance, to study the neural correlations or the causal role of neural activity in choices, the reward pay-off of available options might not be fixed. The experimenter might want to manipulate the probability of reward delivery (eg. Sugrue et al., 2004), the amount of reward delivered (eg. Lau and Glimcher, 2008) or/and the timing of reward delivery (eg. Nomoto et al., 2010). This approach can unravel the neural computations which occur during value transformations of decision variables, that link reward experience to action (Sugrue et al., 2005).

The expected value of a choice is calculated as the product of magnitude (amount) of reward and reward probability (Rolls et al., 2008). If the reward probability of a given action is less than unity then the animal experiences, and has to deal with, a degree of outcome uncertainty, also named *risk*. Neural correlates of risk have been identified - prefrontal cortical areas, striatum and midbrain dopaminergic projections have been previously linked to the processing of this type of uncertainty (Critchley et al., 2001a; Fiorillo et al., 2003; Hsu et al., 2005; O'Neill and Schultz, 2013; Schultz et al., 2000, 2011). While risk is normally measured across outcomes observed over multiple trials, the scope of the work presented hereon will be focused on decision uncertainty, a single-trial estimate on the basis of the decision variables of the current trial (Kepecs, 2013).

How to build a decision and compute confidence

A number of computational strategies have been used to model perceptual decision-making under uncertainty and its neural correlates (Dayan and Abbott, 2005). While being different in the power of predictions and explanations that they can achieve (Churchland et al., 2011; Drugowitsch and Pouget, 2012) they provide a common strategy to

the study of decisions. They incorporate behavioural and neural data and link hidden, internal variables driving behaviour to external, observable variables in a quantitative manner. While it is not the purpose of this dissertation to commit to a particular proposed model, it is useful to mention those that have particularly influenced the way we conducted the research explained here.

Signal detection theory

Signal-Detection-Theory (SDT) is a statistical technique that prescribes a process to convert observations of noisy evidence into a categorical choice, and extract a confidence estimate from this process. For more than 45 years it has allowed psychologists to infer from behavior properties underlying sensory representations (Green and Swets, 1966) and it has provided a strong basis for probing the neural mechanisms that underlie perception and categorization under uncertainty (eg. Salzman & Newsome, 1994). From SDT it is possible to use Receiver Operant Characteristic analysis (ROC), and calculate a measure called area under the ROC curve, to account for the performance of a binary classifier. This approach allows to infer how well a ideal observer could distinguish between two distributions of, say, trial-to-trial firing rates, and classify the signal identity.

In the context of SDT in a categorical decision samples from two noisy distributions are compared and related to each other. Establishing a decision rule to which this comparison should obey to, allows for a decision. The absolute Euclidean difference between the two samples will be a quantitative measure of confidence. This measure of confidence arises from whatever process comparing two distributions, and wherever it occurs. It could be used to analyze any comparison between two distributions of behavioural events or neural firing rates. If the comparison is made between samples of firing rates drawn from two stimuli distributions, this process can tell which stimuli is present, and with what degree of confidence. If it's made between a sensory sample and a sample drawn from a noisy decision rule distribution, this computes which decision to make and the degree of confidence in that

decision. If it's made between firing rates sampled from two different outcome distributions, (e.g. error and correct trials), it can identify which outcome is more likely to be expected, and the certainty of that estimation.

Accumulator models

Accumulator models are a second class of dynamic models of decision-making used to compute decisions under uncertainty and compute confidence estimates (Beck et al., 2008; Gold and Shadlen, 2007; Sugrue et al., 2005; Vickers and Packer, 1982). These models are part of sequential analysis, a natural extension to SDT that accommodates multiple pieces of evidence observed over time. In accumulator models, evidence in favor of a given decision is accumulated over time until it crosses a threshold level and decision is made. In perceptual decisions, neural activity that represents sensory stimuli is used as evidence and so is any neural activity that represents value properties of the different options. The total combined evidence forms a conceptual entity named decision variable, which is “interpreted” by the decision rule to produce a choice. Much effort has been made to describe the neural correlates of this decision variable using accumulator models to understand where and how is it formed (Ding and Gold, 2013; Gold and Shadlen, 2007; Mazurek, 2003; Roitman and Shadlen, 2002), and more recently, how is it related to confidence (Fetsch et al., 2014; Kiani and Shadlen, 2009; De Martino et al., 2013).

Recurrent neuronal circuits

The architecture of recurrent circuits can be used to better understand how could a neural circuitry implement a decision (for a review, Wang, 2008). Generally, a circuit of neuronal populations composed of three distinct layers could serve as the central core for decisions, which incorporate external stimuli, evaluate actions and take decisions. An initial layer of neurons would be a first rail for representing sensory variables used in decisions. Projections from these neurons to an action valuation layer would bias the action to be implemented. Downstream

from this one layer of neurons would be responsible for actually assuming the decision. A circuit comprised of sensory cortices and different regions of basal ganglia could assume the roles of the different layers and be a neural implementation of this process. (Ding and Gold, 2013)

This framework takes into account more precise physiological parameters like the balance between recurrent excitation and feedback inhibition to instantiate attractor states for forming categorical choices and also long transients for gradually accumulating evidence in favor or against alternative options. By adding reward-dependent synaptic plasticity this circuitry can learn to produce adaptive choice behavior. Moreover, recurrent networks could be implemented to derive decisions about confidence estimates, by adding a second recurrent network, fed by the decision-making network (Insabato et al., 2010a).

Neural mechanisms of decision-making

Having in mind the former computational approaches we will proceed by mentioning a foundational body of work, successful in establishing neural basis for decision-making. Next we will introduce some of the research already done in perceptual decision-making, using rats or mice, a line of research that might be the future direction of the field.

The random-dot motion task: Neural circuitry of decision-making in non-human primates

A significant bulk of knowledge on the neural basis of perceptual decision-making has been collected by studying monkeys performing a visual discrimination task - the random-dot motion task (RDM).

In this task a head-fixed monkey, has to decide between two possible (opposite) directions of dots randomly moving in a computer screen. Task difficulty is controlled by varying the percentage of coherently moving dots. On high-coherence trials, the majority of dots move in the

same direction, making it easy to decide the correct global motion direction. On low-coherence trials, only a small percentage of dots move in the same direction, while the other dots move randomly, making the direction decision more difficult. The monkey's movements are restrained and the direction decision is typically indicated with an eye movement, a saccade. In this task response times vary with the difficulty of the decision: responding quickly yields lower accuracy, whereas taking longer to respond corresponds to higher-accuracy decisions (Palmer et al., 2005).

Extensive physiological and microstimulation studies have shown that direction-sensitive neurons in the middle temporal (MT) area of the visual cortex encode motion stimulus (Britten et al., 1992, 1993; Newsome et al., 1989), thus encoding the sensory information relevant for this task. The decision process itself occurs downstream of MT. The final preparation of eye movements involves the superior colliculus (SC) (Horwitz and Newsome, 1999, 2001) but the decision variable, which mediates the transformation of accumulated evidence to a binary choice, is represented midway between MT and SC. This is visible in neural activity from the parietal cortex, area LIP (Kiani et al., 2008; Shadlen and Newsome, 1996, 2001). Related to accumulation of evidence, LIP neuronal activity also correlates with reaction time (Roitman and Shadlen, 2002).

LIP has been implicated in other high-order processes involved in the selection of saccade targets, like representation of bias, reward, expected value, and elapsed time (Dorris and Glimcher, 2004; Leon and Shadlen, 2003; Platt and Glimcher, 1999; Sugrue et al., 2004). Most relevant to the subject proposed in the work presented here, LIP was implicated in decision confidence signals, and its behavioural expression (Fetsch et al., 2014; Kiani and Shadlen, 2009). So the same sensory neural signals in LIP reflect choice, reaction time, and confidence in a decision in monkeys performing the RDM task.

Perceptual decision making in rodents

Inspired by the approaches and results obtained by studying non-human primates in tasks like the RDM, the last decade has seen an increasing amount of scientific advances in perceptual decision-making by experiments involving rats and mice (for a review, Carandini & Churchland, 2013). The rodent brain shares the most fundamental design principals with those of humans, and other primates. Rodents are able to perform simple psychophysical tasks like the ones primates normally perform and, opposite to primate research, data can be collected from a multitude of individuals, with less financial (and ethical) cost (Abbott, 2010). Adding to this is the promise of circuit targeting and manipulation (eg. Cohen et al., 2012; Kvitsiani et al., 2013; Tai et al., 2012; Znamenskiy and Zador, 2013), advantages that can be obtained due to genetic and molecular tools that are more easily implemented in these models.

The odor-guided two-alternative forced-choice task

Rodents can perform psychophysical tasks in response to a plethora of sensory modalities: olfaction (eg. Ranade & Mainen, 2009), audition (eg. Otazu, Tai, Yang, & Zador, 2009), vision (eg. Busse et al., 2011), somatosensory (reviewed in Diamond & Arabzadeh, 2013), or even combining multiple modalities (Raposo et al., 2012). Given the subject of this dissertation, let us focus on olfactory-guided decision making.

One of the most prolific lines of research so far has come from the use of a two-alternative forced choice tasks scheme, where rats have to categorize binary choices in response to olfactory stimuli. In this task a rat pokes his nose into a central “poke” where odors are delivered. After sniffing the odor, it is then free to choose to move towards one of two side pokes, where reward is available, only for correct choices. Rats are psychophysically challenged by mixtures of two odors, with varying contrasts. In this task the mixture contrasts serve the same purpose as the coherence of moving dots in the RDM, such that on average higher accuracy is achieved in response to higher mixture contrasts.

A first behaviour breakthrough that came from animals performing this task was that, contrary to the RDM, response time did not vary with task difficulty, or in other words, speed did not trade with accuracy (Uchida and Mainen, 2003). This might imply that, contrary to the RDM, evidence is not being accumulated over time in this perceptual decision and it could be concluded that rodents do not integrate evidence to commit to decisions, as humans and other primates. So the processes underlying perceptual decision-making could not be generalized across these species. But that is not the case as was later found in rodents performing a task that requires integration (Brunton et al., 2013). What seems to be happening is that the sensory aspect of this task is not limiting to the animals, so although rats were making mistakes, especially for lower contrast mixtures, uncertainty does not arise at the sensory level. It was later suggested that uncertainty in odor category decisions arises from noise sources that fluctuate slowly, from trial-to-trial, rather than rapidly within trials (Zariwala et al., 2013). This view is important for the work presented here.

Establishing a neural circuit for odor categorization

Rodents are very good at smelling and a big portion of their brain is dedicated to processing olfactory cues. The study of the early events of olfactory processing is a wide field on its own (reviewed in Murthy, 2011; Uchida, Poo, & Haddad, 2014; Wilson & Mainen, 2006) and much is already well understood from studies involving non-behaving animal data. The first relay for processing olfactory information is the olfactory bulb (OB), to where olfactory receptor neurons project forming stereotyped structures called glomeruli. These regions encode the identity of odor presented to the rats in the central odor port (Uchida and Mainen, 2003). The firing of OB projection neurons is well locked to the onset of sniffing (Cury and Uchida, 2010) and these projections reach multiple regions in the brain, predominantly the piriform cortex (PC), olfactory tubercle and cortical amygdala (Sosulski et al., 2011). In the PC, single units or neuronal ensembles were found to have correlated activity with multiple aspects of the odor mixtures presented to the

animal, but had a low average correlation between their firing rates and subjects' choices, or expected outcome (Miura et al., 2012a). So, like MT in the RDM task, the piriform cortex appears to represent the sensory information required for task performance. But its activity is not correlated with decision in the odor categorization task. Also, by monitoring the activity from a small subset of PC neurons, a simple decoder based on firing rates could extract more than enough information in a single sniff cycle to account for the behavioral accuracy of rats in the odor categorization task. This adds to the notion that the noise source, which limits performance in this task, lies downstream of the sensory layer of this decision-making circuitry.

In a different set of studies the role of superior colliculus (SC) in the odor-guided categorization task was investigated. Neurons in rats' SC encoded choice direction while the animals were moving towards the side choice ports and inactivation of this region severely disrupted the choice behavior (Felsen and Mainen, 2008). This has broadened the spectrum of orientation-dependent actions with which the SC is involved, away from the “primate” saccade. Furthermore, in a version of the task where animals were forced to remain longer in the central port the activity of SC neurons could often predict the upcoming choice far in advance of movement initiation. Also, choice-predictive neurons were jointly modulated by discrimination difficulty (odor contrast) and choice outcome (error vs. correct trials), demonstrating that the critical sensory information necessary for constructing higher-order decision variables can be carried in the SC along with the choice itself (Felsen and Mainen, 2012). In this same version of the task, performed by mice, activity of neurons from a brain stem region (pedunculopontine tegmental nucleus), was also correlated with choice direction and outcome (Thompson and Felsen, 2013). These two previous results indicate that properties of the decision variable are still reflected in the activity of neurons from regions closer to the motor output responsible for movement. But given that animals do not improve performance by sampling odors for longer times, it's likely that the decision had been committed already while animals were forced to remain still in the odor

port. As in the RDM, it's likely that the decision starts to take shape upstream from the SC.

In a similar task, where rats have to perceive auditory stimulus to guide decisions, optogenetic manipulation of auditory cortex neurons which project to auditory striatum, biased animals' choices (Znamenskiy and Zador, 2013). This bias was highly specific, towards or against the direction predicted by the cortical tonotopy, depending on whether the neurons were activated or inhibited. Given the ubiquity of corticostriatal projections in cortex, the authors suggested that these might provide a general mechanism for control of motor decisions by sensory context. In a version of the odor guided categorization task that included a reward value manipulation, neurons from the dorsomedial striatum (and ventral striatum, to a smaller extent) were shown to play a critical role in net value-dependent regulation of response vigor (Wang et al., 2013). Nevertheless, the neuronal activity correlated with this effect was observed before trial initiation and no perceptual-related bias of choice was reported, leaving open the question of whether this area plays a role in biasing decisions with odor specificity and whether trial-by-trial decision uncertainty in this task could arise from noise in cortico-striatal associations.

Neural correlates of decision confidence and uncertainty in the orbitofrontal cortex

The role of the orbitofrontal cortex (OFC) was also investigated in the olfactory-guided categorization task. In this task, neurons in the OFC were found to encode information about presented stimuli, but not about choice, in the period of time before a decision was made (Feierstein et al., 2006). In the same study OFC neurons encoded choice direction when the animal was moving towards the choice port. 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, which was shown in the context of reward-based decision making (Balleine et al., 2011; Burke et al., 2008; Fellows, 2011; Morrison

and Salzman, 2011; Padoa-Schioppa and Assad, 2008; Padoa-Schioppa and Cai, 2011; Roesch et al., 2006; Schoenbaum et al., 2011; Schultz et al., 2000; Takahashi et al., 2013; Wallis, 2012). OFC activity was also found to be encoding decision confidence in rats performing the odor categorization task (Kepcs et al., 2008). This finding was foundational for the work presented here, so let us describe it in detail.

A simple modification of the odor categorization task allowed for a better focus on neural representations of uncertainty. In this task reward delivery was delayed, so the animal had to spend time (0.3-2s) in the choice port, anticipating the trial outcome. This was called the outcome anticipation period. Everything else remained similar - the different odor mixture contrasts allowed for challenging choices and deterministic reward contingencies enforced the decision rule. During the outcome anticipation period 24% of the recorded OFC units were firing as a function of odor mixture and outcome forming a “x”-pattern characteristic for uncertainty signaling – on average they were less active for correct trials, in a difficulty dependent manner (**Figure i E-H**). Firing rates were lower for correct trials for easy stimuli (100% odorA : 0% odor B and 0% odorA : 100% odor B mixtures) than for correct trials for stimuli closer to the 50A:50B category boundary. The opposite was observed for error trials. So, the firing rate of these cells was inversely correlated with accuracy, with higher firing rates predicting chance performance. These were called the “uncertainty” population of neurons. In 19% of the recorded units, the opposite pattern was observed. These were called “confidence” cells. In our framework, confidence is simply the opposite of uncertainty.

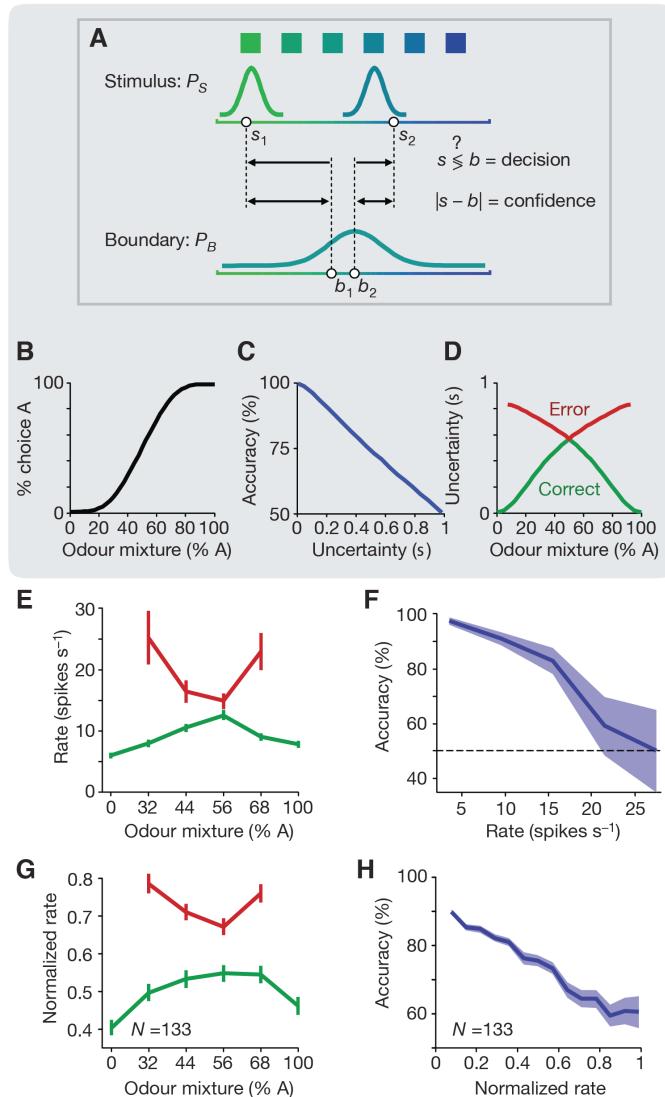


Figure i - Confidence estimation in a decision model and by OFC neurons.

A, Schematic of a model for category decisions. Each odour mixture stimulus, as well as the memory for the category boundary, is encoded as a distribution of values. In each trial a stimulus, s_i , and memory of the boundary, b_i , are drawn from their respective distributions. A choice is calculated by comparing the two samples (s_i, b_i), and a confidence value is estimated by calculating their distance ($|s_i - b_i|$). Incorrect choices result from noise, represented in the model by the width of the stimulus and category boundary distributions.

B, Example psychometric function of the model, replicating the high choice accuracy of rats for pure odours and decreased accuracy for mixtures near the imposed category boundary.

C, Mean accuracy of model choices as a function of decision uncertainty. The uncertainty estimate, s , is transformed from the distance between the stimulus and boundary

samples. **D**, Mean decision uncertainty estimates generated by the model as a function of stimulus and trial outcome. Note that the model (or a subject) has access only to a stimulus sample and not the stimulus type (for example, 56/44). **E**, Firing rate of an example neuron during the outcome anticipation period as a function of odour stimulus and trial outcome. Error bars are s.e.m. across trials. **F**, Mean choice accuracy as a function of the firing rate for the same unit in E. Firing rates were binned and the mean accuracy was calculated for each range of firing rates. Error bars represent standard errors based on the binomial distribution of outcomes. **G**, Mean normalized firing rate of negative outcome selective population (negative outcome preference index across trials with all stimuli pooled at $P < 0.05$, permutation test) during the anticipation period. **H**, Mean accuracy as a function of the firing rate for the same neuron population as in G. Firing rates were binned for individual neurons and the mean accuracy was calculated for each range of firing rates. These curves were normalized to a maximal firing rate of 1 and averaged. Error bars represent s.e.m. across neurons.

- Adapted from Kepcs et al 2008.

These patterns of activity were predicted by an SDT-based model that derived choices and confidence estimates for the odor categorization task, **(Figure i A-D)**. Also a race-model, a type of accumulator model, could predict the same pattern of activity, which contributed to the notion that confidence estimation is a generalized property inherent to the decision making process based on sensory stimuli. Additionally,

activity of these cells was not explained by the history of previous rewards, choices or outcomes, suggesting that these signals were related to a decision just made.

This was the first evidence for decision confidence neural correlates in animals. Subsequent research, in non-human primates, within the scope of visual guided decision making, have found neuronal correlates of decision confidence in cortical areas such as the LIP (Kiani and Shadlen, 2009), the supplementary eye field (Middlebrooks and Sommer, 2012), the pre-motor cortex (Martinez-Garcia et al, 2014) and the pulvinar nucleus of the thalamus (Komura et al., 2013).

In the Kepecs study the authors trained rats to perform one other version of the task named the “reinitiation” task. Here, the reward delivery delay was even longer, varying from 2 seconds up to 8 seconds. In incorrect trials, an auditory tone played 8 seconds after choice port entry signaled the error and was followed by a time-out punishment. Animals were allowed to leave the choice port before the 8 seconds had passed and the task design made it so that upon initiating a new trial they could expect a higher contrast (easier) stimuli to be delivered.

The fraction of trials in which animals decided to reinitiate a trial followed the same “x”-pattern as observed from firing rate of uncertainty cells in the original task– they reinitiated a trial more often when expecting the trial to be incorrect, after a high contrast mixture. This was indicative that animals could use confidence information behaviorally. However, no electrophysiological data was acquired when animals were performing the reinitiation task and it was left to know how did the activity from OFC uncertainty (and confidence) cells correlated with the behavioural usage of confidence.

One other caveat related to this study comes from the previous history of research in the OFC, which mostly linked activity of this area to encoding properties of reward value (Wallis, 2007). Due to the two-alternative forced choice task design the data is consistent with either a representation of outcome probability or outcome uncertainty signals.

Either way, the computation of outcome probability must incorporate an estimate of decision uncertainty, and this estimate is thought to be important for computation of subjective expected value (Mainen and Kepecs, 2009). Since the value of reward was kept constant, it is not clear how does the uncertainty signal varies with magnitude of expected reward.

Evaluation, or performance monitoring, is necessary to analyze the efficacy or optimality of a decision with respect to its particular goals (Shadlen 2007). OFC is particularly important for reward-based behaviors when values are inferred, for instance using model-based reinforcement learning algorithms (Daw and Doya, 2006; Jones et al., 2012; Wilson et al., 2014). Sharing a similarly important role in evaluation of performance is the ventral-striatum (VS) (eg. 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 likely that activity in VS neurons also correlates with decision confidence, (Daniel and Pollmann, 2012a; Hebart et al., 2014a), but whether these neurons can predict a behavioral report of confidence is also not known. A cortico-striatal circuit involving OFC and VS could be relevant to evaluate decisions and optimize actions taking into account confidence estimates.

Behavioral use of confidence

When studying human decision making one can simply ask a subject for a semantic confidence report or raking about the decision made. While these might (or might not) be an accurate report of decision confidence, it is a strategy that cannot be followed in animal studies. To overcome this “communication” issue there is a realm of behavioural tasks designed to study confidence in animals. They can be subdivided in three categories: uncertain option tasks, opt-out tasks and post-wagering tasks (reviewed in Kepecs & Mainen, 2012).

Generally, in uncertain option tasks (eg. Smith, Shields, Schull, & Washburn, 1997) animals can choose to perform a two-alternative perceptual categorization or choose an uncertain option. In these tasks the reward payoff matrix is as follows - a correct categorization leads to reward, choosing the uncertain option leads to a smaller reward amount and wrong categorizations lead to absence of reward, or time-out punishment. In these tasks animals tend to choose more often the uncertain option whenever the stimulus presented is close to the category boundary. But a caveat of these tasks is that they can be viewed as a three-choice decision, which can be solved simply by learning appropriate stimulus-response categories without necessitating confidence estimates.

Opt-out tasks are a variation of uncertain option tasks. They have been implemented in the RDM task scheme and allowed for the discovery of neural correlates of decision confidence in the parietal cortex of monkeys (Kiani and Shadlen, 2009). In Kiani's study (as in all other RDM tasks) animals had to make a binary choice dependent on the motion of visual stimuli. On a subset of trials, after stimulus presentation, a third 'opt out' choice was presented for which monkeys received a smaller but guaranteed reward.

The authors found that the frequency of choosing the 'opt out' choice increased with increasing stimulus difficulty and with shorter stimulus sampling. Moreover, performance on trials in which they declined to opt out was better than when they were forced to perform the discrimination. A problem with opt-out tasks, which is shared with uncertain option tasks, is that the task design does not allow for confidence report and perceptual guided decision to occur in the same trial. So confidence measures related to decisions just made cannot be obtained. Furthermore, confounding factors, such as attention or motivation, can be used to explain why it is that when animals choose not to opt out they perform better (Kepes and Mainen, 2012).

An ideal confidence-reporting task requires a report of choice and the confidence associated with that choice in the same trial. Post-decision

wagering tasks can accomplish that, given that after the choice is made, confidence is assessed by asking a subject to place a bet on his choice (Persaud et al., 2007). The reinitiation task from Kepecs was similar to a post-decision wager because after the odor-guided decision, rats would have to make a new decision in whether to stay and risk no reward with timeout punishment, or leave and start a new, potentially easier trial. But it still carried a problem - in each trial only a single bit of information was gained about decision confidence (stay or leave). A graded report of confidence might be preferable over a binary report, for instance, so that single trial neural activity can be correlated with single trial confidence-driven behavior.

AIMS

Neural and behavioral correlates of decision confidence

In this dissertation we will aim at the following:

Design a post-wager confidence report suited for rodents and interpret the behavior using theoretical frameworks.

Evaluate whether OFC is required for confidence reports.

Explore OFC and VS neural activity in the computation of decision confidence and behavior adaptations tuned by confidence

Understand the effect of reward magnitude in post-decision confidence judgments

ORBITOFRONTAL CORTEX IS REQUIRED FOR OPTIMAL WAITING BASED ON DECISION CONFIDENCE

Results published: Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence, Lak A, Costa GM, Romberg E, Koulakov AA, Mainen ZF, Kepcs A. *Neuron*. 2014 Sep 17. pii: S0896-6273(14)00740-5. doi: 10.1016/

Author contributions

Kepcs A., Lak A and Costa GM designed the studies. Lak A and Costa GM ran the experiments, and analyzed the data. Koulakov AA and Kepcs A designed the model and Lak A fitted the results. Kepcs A, Lak A and Mainen ZF wrote the manuscript.

SUMMARY

Confidence judgments are a central example of metacognition—knowledge about one’s own cognitive processes. According to this metacognitive view, confidence reports are generated by a second-order monitoring process based on the quality of internal representations about beliefs. Although neural correlates of decision confidence have been recently identified in humans and other animals, it is not well understood whether there are brain areas specifically important for confidence monitoring. To address this issue we designed a post-decision temporal wagering task in which rats expressed choice confidence by the amount of time they were willing to wait for rewards. We found that orbitofrontal cortex inactivation disrupts waiting-based confidence reports without affecting decision accuracy. Furthermore, we show that a normative model can quantitatively account for waiting times based on the computation of decision confidence. These results establish an anatomical locus for a metacognitive report, confidence judgment, distinct from the processes required for perceptual decisions.

INTRODUCTION

If you are asked to report your confidence in a decision—how certain are you that you made the correct choice—you can readily answer. But what is the neural basis for this ability? Early behavioral studies considered confidence judgments as a type of metacognitive process related to self-awareness. These studies established that several species besides humans are capable of confidence judgments but that some, such as rats, may not be (Flavell, 1979; Hampton, 2001; Metcalfe, 2008; Smith et al., 2003). Against this backdrop of behavioral results, a recent line of studies identified single neuron correlates of decision confidence across species, in the brains of rats and monkeys (Kepecs et al., 2008; Kiani and Shadlen, 2009; Komura et al., 2013; Middlebrooks and Sommer, 2012), as well as functional correlates in humans (Fleming and Dolan, 2010;

Lau and Passingham, 2006; De Martino et al., 2013; Rolls et al., 2010a; Yokoyama et al., 2010). However, it is still not well understood where and how choice confidence is computed or how it is made accessible to an overt behavioral report. These issues are particularly interesting because they relate to the definition of metacognition and awareness.

A mechanistic interpretation of metacognitive theories implies that a second-order brain circuit reads first-order representations of a separate circuit and transforms them into a second-order representation, such as a decision variable for confidence (Insabato et al., 2010a; Kepcs et al., 2008; Kiani and Shadlen, 2009; Komura et al., 2013; Ma et al., 2006; Middlebrooks and Sommer, 2012). The representation of decision confidence in specific brain regions implies that lesions of such brain areas might affect the behavioral manifestation of decision confidence without changing other aspects of the choice behavior. In contrast, theoretical studies suggest that because confidence estimation is central to statistical inference, it ought to play a fundamental role in probabilistic or Bayesian neural computations of all kinds (Ma et al., 2006; Moreno-Bote, 2010; Rao, 2010; Zemel et al., 1998). This view suggests that the computations of choice and confidence are mixed within the same neural circuits and hence representations of confidence might not be explicit or anatomically segregated(Higham, 2007). Consistent with these ideas, data from primates show that neurons in parietal cortex that represent a perceptual decision also encode the confidence associated with that decision(Kiani and Shadlen, 2009).

Here we pursued the hypothesis that orbitofrontal cortex (OFC) is causally required for confidence reporting independent of perceptual decision-making. This hypothesis was based on two lines of evidence. First, previously we found that rat OFC contains an explicit representation of decision confidence (Kepcs et al., 2008). Second, OFC has been implicated in goal-directed or intentional decisions that require the evaluation of predicted outcomes (Jones et al., 2012; Kennerley et al., 2011; Morrison et al., 2011; Padoa-Schioppa and Assad, 2008; Rolls and Grabenhorst, 2008; Schoenbaum et al., 2009; Wallis, 2007).

Because reporting confidence requires performing an action on the basis of a predicted outcome, an intact OFC may be required for adaptive adjustment of the behavior according to decision confidence. At the same time OFC is probably not involved in most perceptual decisions.

Studying confidence reports in animals requires a clear behavioral readout of confidence. Gambling on the outcome of a decision generates an observable wager that can quantitatively index confidence (Middlebrooks and Sommer, 2012; Persaud et al., 2007). Appropriate wagering requires an evaluation of decision confidence that can be distinguished from random betting using a computational approach (Fleming and Dolan, 2010; Kepcs and Mainen, 2012; Kepcs et al., 2008; De Martino et al., 2013). Therefore to evaluate whether OFC is required for confidence reports of perceptual decisions, we designed a gambling task for rats with continuous wagers based on their willingness to wait for delayed rewards, interpreted the wagers within a theoretical framework for statistical confidence and used inactivation methods to probe the role of OFC in waiting-based confidence judgments.

RESULTS

A post-decision wagering task

To study confidence in perceptual decisions we used an extensively studied odor categorization task that allowed us to systematically vary the difficulty and hence confidence in a decision (Kepcs et al., 2008; Uchida and Mainen, 2003). Upon entry into a central odor port, rats ($n=10$) received an olfactory stimulus (binary mixture of 2-octanol stereoisomers) and responded to the left or right choice ports based on the dominant odor component (**Figure 1.1A**, See Experimental Procedures). Trials with different odor-mixture ratios (20:80, 40:60, 43:57 and 50:50 mixtures and their conjugates, 80:20, etc.) were randomly interleaved. Rats achieved high performance for easy stimuli (larger mixture ratios), but were challenged by more difficult

discriminations (**Figure 1.1B**). The perceptual accuracy was stable across several sessions of testing (**Figure 1.1B**). As previously reported (Uchida and Mainen, 2003; Zariwala et al., 2013) reaction times, as measured by the duration animals were sampling the odor before moving to the choice port, showed little sensitivity to odor-mixture ratio (**Figure 1.1C** and **Figure 1.S1**).

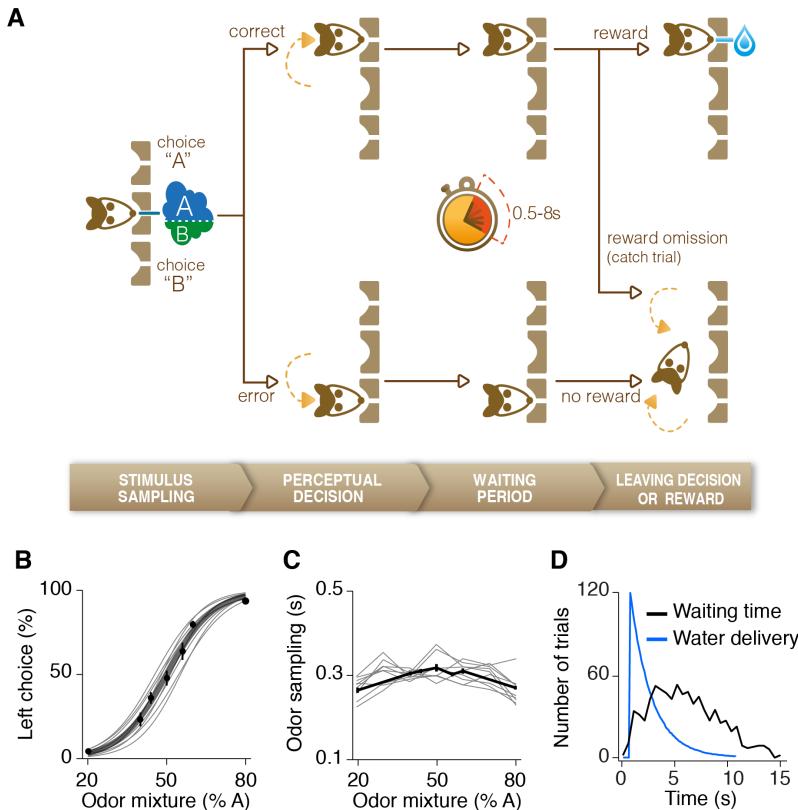


Figure 1.1- Post-decision wagering task using temporal wagers.

(A) Schematic of the behavioral paradigm. To start a trial, rats entered the central odor port and after a pseudorandom delay of 0.2–0.5 s a mixture of odors was delivered. Rats responded by moving to the left or right choice port, where a drop of water was delivered after a 0.5–8 s waiting period for correct decision (exponentially distributed with a decay constant of 1.5 after a 0.5 s offset and 8 s maximum). In a small fraction of correct choice trials, water rewards were omitted. Trials of different odor-mixture ratios were randomly interleaved independent of rats' performance in the previous trials. While waiting for reward, animals were required to keep their snouts inside the choice port, which was continuously monitored using infrared photo-beams. Failure to break the photo-beam resulted in error. **(B)** Behavioral performance and psychometric function of an example rat. Each thin line represents logistic fit (see Experimental Procedures) to the behavioral data collected in a single test session. Dots represent behavioral performance averaged across all trials of all test sessions. Thick gray line represents logistic fit to the average performance data shown with black dots. Error bars in all panels represent \pm s.e.m. across trials. **(C)** Odor sampling duration (the duration animals were sampling the odor before moving to the choice port) as a function of odor mixture contrast in an example rat. Thin lines represent odor sampling duration in each of test sessions. Thick line represents the data averaged across all trials of all test sessions. **(D)** The timing of reward delivery (blue, see Experimental Procedures) and the distribution of waiting times at the reward ports of all test sessions for one example rat (black). Waiting times were measured for all the error trials and fraction of correct trials (i.e. reward omission trials).

An ideal confidence-reporting task requires a report of choice and the confidence associated with that choice in the same trial (Kepcs and Mainen, 2012; Middlebrooks and Sommer, 2012). Lacking this, it is difficult to rule out alternative mechanisms, a limitation of opt-out tasks (Kepcs, 2013). In addition, a continuous, rather than discrete, behavioral measure of confidence might enable stronger inferences about the underlying mechanisms (Kepcs and Mainen, 2012; Schurger and Sher, 2008). To allow rats to wager on the likelihood that their decision was correct, we delayed reward delivery and measured the time animals were willing to wait at the choice ports (**Figure 1.1A,D**). Reward delay was drawn from an exponential distribution (decay constant, $\tau = 1.5$, see Experimental Procedures) to generate a relatively constant level of reward expectancy over a range of delays (i.e. flat hazard rate), (Janssen and Shadlen, 2005; Zariwala et al., 2013). Incorrect choices were not explicitly signaled and hence rats eventually left the choice ports to initiate a new trial. To measure waiting time (WT) for correct choices we introduced a small fraction of catch trials (10-15%) for which rewards were omitted. Therefore, in this novel post-decision wagering paradigm, each trial resulted in a binary choice as well as a graded wager, WT, for all incorrect trials and a fraction of correct trials.

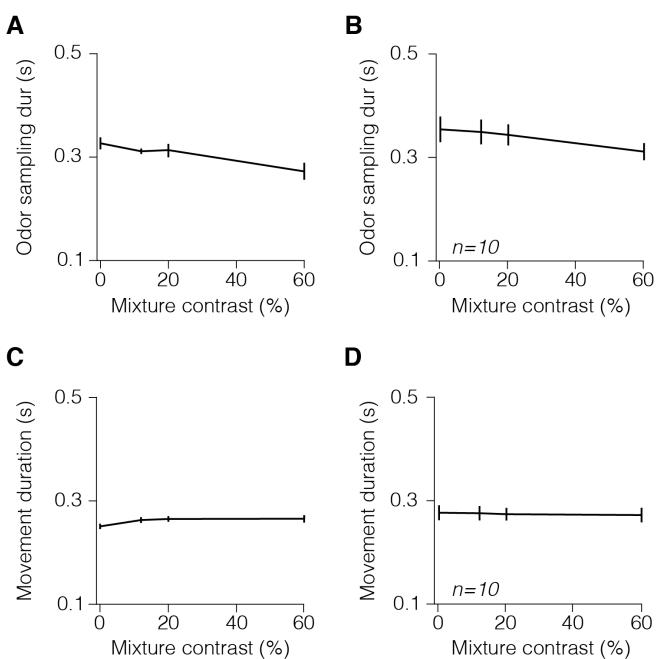


Figure 1.S1- Response times

(A) Mean odor sampling duration (defined as the interval between odor valve opening time and the time point on which rat leaves the odor port) for the example rat as a function of odor mixture contrast. For all panels of the figure, error bars are \pm s.e.m across trials or across rats. **(B)** Mean odor sampling duration as a function of odor mixture contrast averaged across ten rats. **(C)** Mean movement duration (defined as the interval between the time of leaving the odor port and the time of entry into a choice port) as a function of odor mixture contrast for the same rat as in (A). **(D)** Mean movement duration as a function of odor mixture contrast averaged across ten rats.

Derivation of optimal confidence-based temporal wagering

To maximize rewards, an ideal observer should wait until the relative expected value of waiting for reward drops below the expected value of leaving. Because reward size is constant from trial-to-trial but depends on being correct, the subjective expected value of staying varies from trial-to-trial with the level of decision confidence. To derive the normative waiting time we assumed that the observer arrives at the reward port with a specific internal expectation about how likely it is to receive the reward, reflecting its decision confidence, which we denoted by the variable C . Assume that the observer has spent time t in the port without receiving a reward. The observer then faces the decision whether to spend the next interval of time, from t to $t+dt$ inside the reward port or leave and initiate a new trial. This decision should be based on the reward hazard function—the probability of getting reward at the next moment given that no reward has been received until now (**Figure 1.2A**). This probability can be computed through Bayes' theorem. Let us denote as W_t the event that waiting until time t was not rewarded, and R the event that reward arrives at the next moment, from t to $t+dt$. The reward expectation (hazard) function can be expressed as the conditional probability $P(R|W_t)$:

$$P(R|W_t) = \frac{P(W_t|R)P(R)}{P(W_t)} . \quad (1)$$

Notice that by definition the probability of waiting without reward given that reward arrives in the next moment, $P(W_t|R)$, is 1. The probability of being rewarded at the next moment, $P(R)$, depends on the subject's estimate of the time of reward delivery. We denote the experimenter-defined temporal distribution of rewards during the anticipation period as $P_{\text{rew}}(t)$, a distribution that was kept fixed during testing (see Figure

1D). Then the probability of getting rewarded at the next moment is given by:

$$P(R) = P_{\text{trial}} \times P_{\text{rew}}(t) \times dt, \quad (2)$$

where P_{trial} is the expectation of being rewarded in the current trial for the given choice. Because we are describing the reasoning of the ideal observer, the expectation to be rewarded should be based on the internal representation of response accuracy, which means that P_{trial} can be associated with the decision confidence, i.e. $P_{\text{trial}} = C$. The probability of waiting until time t without reward can be evaluated as $1 - P(\bar{W})$, i.e. as one minus the probability of being rewarded during that time.

$$P(W) = 1 - C \int_0^t P_{\text{rew}}(t') dt' \quad (3)$$

From these equations we can compute the probability of being rewarded within time interval from t to $t + dt$ under the condition that the reward was not delivered before that

(4)

$$\frac{P(R|W_t)}{dt} \equiv \rho(t) = \frac{C \cdot P_{\text{rew}}(t)}{1 - C \int_0^t P_{\text{rew}}(t') dt'}$$

Here $\rho(t)$ is the rate of reward expected by the observer within the next time interval, which is the reward expectation per unit time. Since, in our experiments, $P_{\text{rew}}(t) = \exp(-t/\tau)/\tau$, we obtain the reward hazard as a function of decision confidence

$$\rho(t) = \frac{1}{\tau} \frac{Ce^{-t/\tau}}{1 - C + Ce^{-t/\tau}} \quad (5)$$

To obtain the optimal waiting time, the rate of reward expectation $\rho(t)$ should be compared to the average reward rate for the session, κ ,

representing the value of leaving or opportunity cost. Indeed, if $\rho(t) < \kappa$, i.e. the expected rate of reward falls below the opportunity cost, the observer should leave the port and initiate a new trial (**Figure 1.2A**).

The optimal waiting time, WT^{opt} , can therefore be obtained from the equation $\rho(WT^{\text{opt}}) = \kappa$, where $\rho(t)$ is given by Equation 5, while κ is a parameter similar across trials within the same block. From this equation, the optimal waiting time is a function of the decision confidence, C (**Figure 1.S2**):

$$WT^{\text{opt}} = \tau \ln\left(\frac{C}{1-C} \frac{1-\kappa\tau}{\kappa\tau}\right) \quad (6)$$

Here C is decision confidence variable from trial to trial, while κ is the opportunity cost (a constant). Opportunity cost is expected to be smaller than $1/\tau$, since otherwise the ideal observer would not have an incentive to go to the reward port. Thus, the product $\kappa\tau$ is less than one. This derivation reveals that WT^{opt} monotonically increases with confidence levels, consistent with intuition (**Figure 1.2B** and **Figure 1.S2**). The equation predicts that when $\kappa > C/\tau$, then WT is zero; meaning that in very low confidence trials or when the opportunity cost is large, it is not worth for the observer to wait inside the reward port. Thus in these cases the animal should abort the trial as quickly as possible.

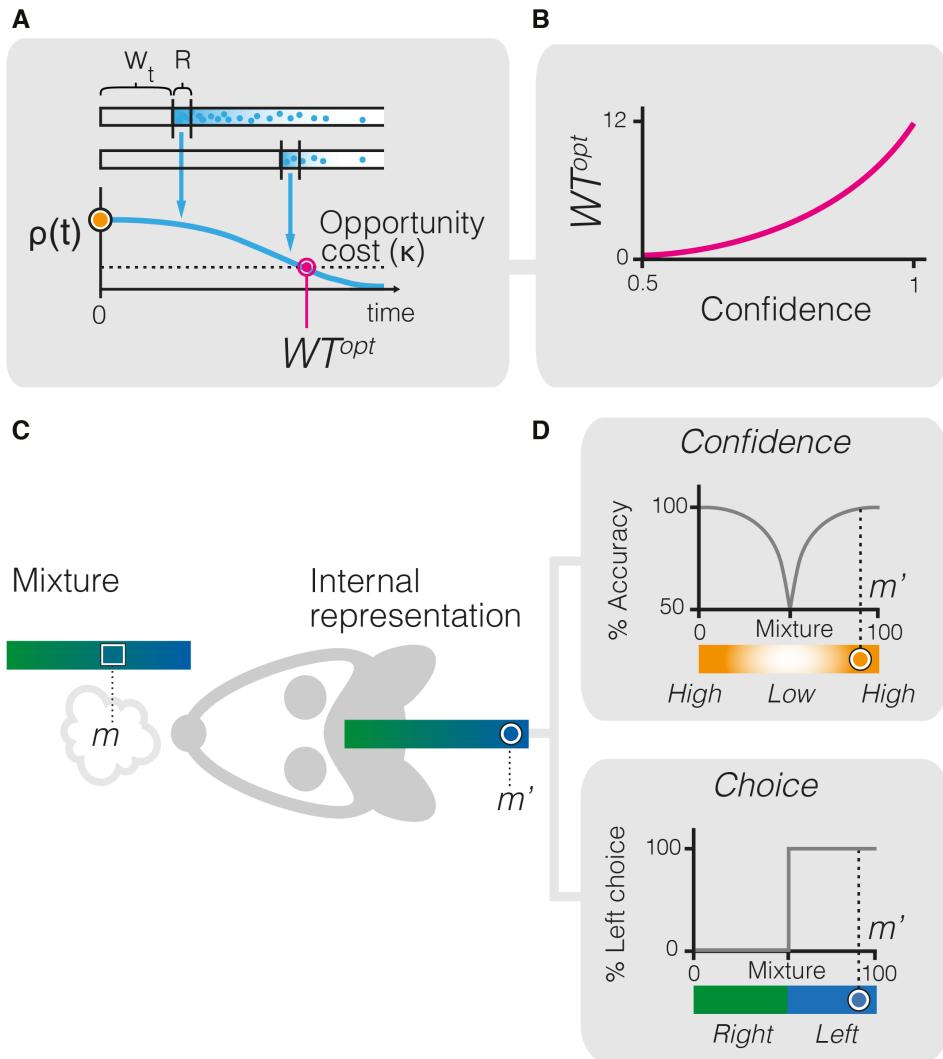


Figure 1.2- A computational framework for estimating waiting time based on decision confidence

(A) The optimal waiting time can be estimated by comparing the rate of reward expectation $\rho(t)$ and the opportunity cost, K . While waiting for a reward, the agent faces the decision each moment. If the expected rate of reward falls below the opportunity cost, the observer should abort the trial and initiate a new trial. The rate of reward expectation $\rho(t)$ is the probability that a reward will arrive at the next moment (denoted as event R), given that the agent did not yet receive reward (denoted as event W). **(B)** The model predicts that WT^{opt} monotonically increases with the level of decision confidence, C . **(C)** In each trial, the stimulus is defined as the percentage of one of the components in the odor mixture (m) and the internal representation of the stimulus (m') is noisy read-out of the external stimulus. **(D)** Choice in each trial is computed by comparing the value of m' and the decision boundary ($b=50\%$), thus a step function of m' . Decision confidence, is a function of the distance between the internal representation of stimulus m' and the decision boundary, as defined by Equation 7.

To model decision confidence, we used a signal detection theory framework where each choice and its associated confidence could be estimated by comparing the sampled stimulus and the decision boundary. We modeled the stimulus as the percentage of one of the

components in the odor mixture henceforth denoted by m and defined a noisy read-out of that as the internal representation of the stimulus m' (**Figure 1.2C**, see Experimental Procedures I). In each trial, the values of m' exceeding the decision boundary ($b=50\%$) result in a response to the right, while the values $m' < 50\%$ produce a left response (**Figure 1.2D**). The distance between the internal representation of stimulus m' and boundary provides an estimate of decision confidence, C (**Figure 1.2D**). Specifically, decision confidence in our approach is defined as the probability of making the correct decision $C = C(m')$ (see Experimental Procedures). It is not difficult to see that for a simple decision task described here, the probability of being correct is

$$C(m') = \frac{1}{2} \left[1 + \text{erf} \left(-\frac{|m' - b|}{\sigma\sqrt{2}} \right) \right] \quad (7)$$

Thus confidence, C , is an internal metric about the probability of choice correctness. Because the internal representation of the stimulus m' varies from trial to trial even if the stimulus mixture m is fixed, response accuracy becomes coupled with decision confidence. Since $C = C(m')$ is an internal variable, it is not available for direct measurement but it could be assessed through the time spent by the observer in the reward port as described above (Equation 6). Notably, these general predictions about decision confidence are not only robust to various forms of stimulus and decision boundary distributions, but could also be derived from other decision frameworks based on Bayes' rule, integration of evidence and attractor models (Kepecs and Mainen, 2012).

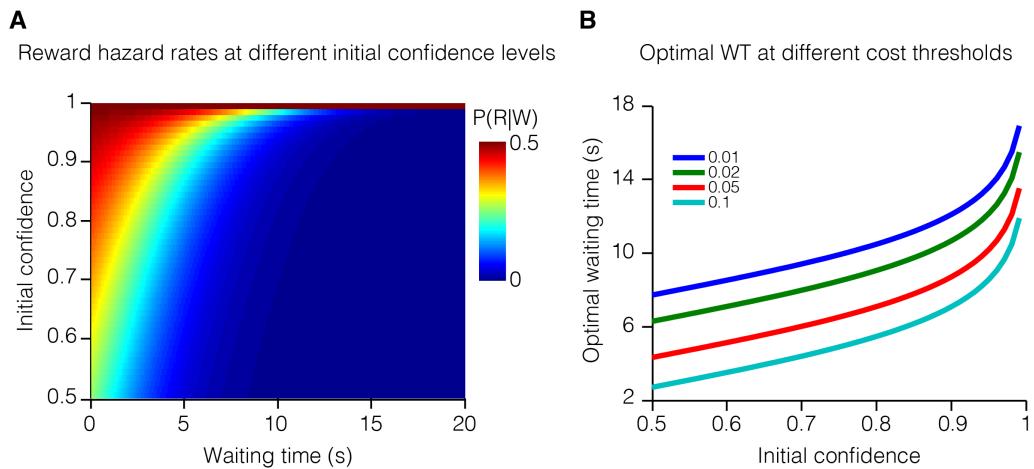


Figure 1.S2 - Ideal observer model for temporal wager

(A) Reward hazard rates (equation (5) in the main text) plotted as a function of waiting time for different levels of decision confidence. For illustration, time constant for the temporal reward distribution was set to 2s. **(B)** Optimal waiting times (WT^{opt} , equation (6) in the main text), predicted by the model as a function of initial decision confidence for different values of opportunity costs, shown with different colors. WT^{opt} is higher when the model encounters lower opportunity costs. For a fixed opportunity cost, the WT^{opt} monotonically increases as a function of decision confidence.

Rats' behavior is consistent with the normative temporal wagering model

We used this computational framework to examine whether rats' WTs could be used as a trial-by-trial proxy of decision confidence. To do so, we fitted our model to rats' behavior (Figure 1.3). Starting from the rat's psychometric curve, we estimated the overall choice uncertainty (standard deviation of the overall sensory and internal noise distribution, σ , see Experimental Procedures I for details of fitting). We then used the estimated σ to calculate the intermediate variable, decision confidence (C), for each trial (Figure 1.3A, middle panel). We then fitted a single free parameter, the opportunity cost, κ , that minimized the difference between rat's and the model's WT distribution (Figure 1.3A, right panel). Although this model fit the mean WTs for each condition well, in order to fit to the full WT distribution we also assumed that rat's estimation of elapsed time carries uncertainty. Specifically, previous studies have shown that the standard deviation of time estimates scales with elapsed time; referred to as 'scalar timing' (Gibbon, 1977; Gibbon et al., 1997; Janssen and Shadlen, 2005). Therefore, for the fitting, the

model's WT distribution was blurred with a normal distribution whose standard deviation was proportional to the elapsed time (Figure 1.3A, right panel, see Experimental Procedures I). As expected from Equation 6, following this fitting, the WTs showed a monotonic relationship with the estimated confidence levels (Figure 2B), demonstrating that WTs in the task could be viewed as a trial-by-trial proxy for decision confidence.

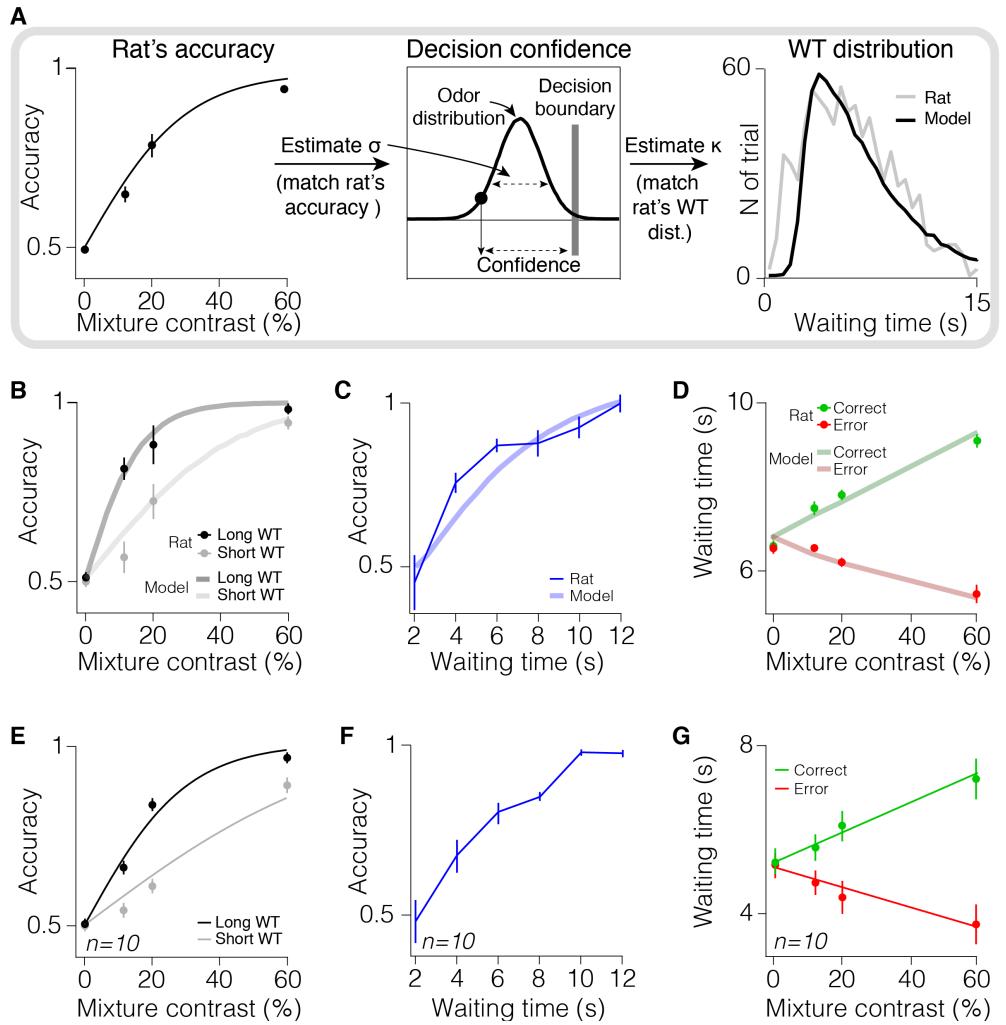


Figure 1.3 - Post-decision waiting time report follows decision confidence.

(A) Fitting the computational model to the behavioral data. Two parameters need to be estimated. First, the standard deviation of the sensory and internal noise distribution (σ), which was used to calculate model's trial by trial choice and confidence. Second, the opportunity cost (κ), which, alongside confidence and reward delay distribution (Equation 6) was used to calculate model's WT. Left and middle panels: estimating the model's noise from rat's psychometric curve. σ was estimated which could minimize the difference between rat's and model's psychometric curves. The estimated σ was used to estimate the confidence associated with a choice in each trial. Next, the opportunity cost, κ , was estimated by minimizing the difference between a rat's and the model's WT distributions. Following the fitting, the model's WT distribution closely overlapped rat's WT distribution. See Experimental Procedures I for details of fitting. **(B-D)** Predictions of the model and behavioral data from example rat. The model produces testable predictions about the relationship between confidence, perceptual accuracy, stimulus difficulty and trial outcome. The predictions of the model closely match the behavioral data. In each panel, thick lines represent the predictions of the model (with parameters optimized to fit rat's accuracy curve and overall WT distribution) and behavioral

data are shown as mean \pm s.e.m. across trials. **(B)** The model predicts that decisions with longer WT have higher accuracy (thick lines). Lines show model's psychometric curves separated based on WT. Dark gray thick line represents long WT (defined as above 70 percentile), light gray thick line represents short WT (shorter than 70 percentile). Dots show rat's perceptual accuracy separated based on WT. Black dots represent long WT trials (defined as above 70 percentile) and gray dots indicate short WT trials (shorter than 70 percentile). Logistic psychometric fits, used for the slope comparison, are not shown. **(C)** In the model, WT predicts choice accuracy (thick line). Consistent with this prediction, rat's decision accuracy increases with longer WT (thin line). **(D)** The model predicts that waiting time varies with stimulus difficulty in opposing directions depending on choice correctness (thick lines; correct: green, error: red) and rat's WTs are consistent with this prediction. Dots show mean WT of the example rat as a function of odor mixture contrast and trial outcome (correct: green, error: red). **(E-G)** As in (B-D) averaged across 10 rats (mean \pm s.e.m. across rats). In (E) black and gray lines represent logistic fit on the accuracy data in long and short WT trials, respectively (see Experimental Procedures). In (G) lines represent linear fits on the rats' WT data.

This model yields specific predictions about how WT, as a proxy for decision confidence, relates to other experimentally controlled and monitored variables. First, decisions in trials with longer WT are expected to have higher accuracy (**Figure 1.3B**, thick lines) for any given stimulus. Consistent with this prediction, when we separated behavioral trials into long and short WT, choice accuracy in trials with intermediate odor mixture contrast showed significant dependency on WT (**Figure 1.3B**, E, $P < 0.05$, Mann-Whitney U-test across trials in 10/10 rats; and $P < 0.05$, Mann-Whitney U-test across rats). The slope of the rats' psychometric functions was also steeper for long WT trials ($P < 0.05$ in 10/10 rats; and $P < 0.001$ across rats, bootstrap test, see Experimental Procedures).

Second, WT is expected to predict choice accuracy (**Figure 1.3C**, thick line). Consistent with this prediction we found that animals' WT-conditioned accuracy function (see Experimental Procedures) monotonically increased with longer WT, ranging from chance level to near-perfect performance (**Figure 1.3C**, F). Third, WT is expected to vary with stimulus difficulty in opposite directions depending on choice correctness (**Figure 1.3D**, thick lines). Indeed, we found that rats' mean WTs varied with stimulus difficulty, and this relationship was opposing for correct and error trials (**Figure 1.3D**, G). For all these predictions, the model with parameters optimized to fit the rats' overall WT distributions (**Figure 1.3A**) showed striking match to their behavioral data, as can be seen in **Figure 1.3B-D**. These properties further established WT as a good trial-by-trial proxy of C , and suggest that it can serve as an implicit report of decision confidence.

We also examined the effects of trial history on the stability and confidence-dependence of WT. We found that the mean WT was stable from the beginning to the end of a session (**Figure 1.S3A, B**, $P > 0.3$, Mann-Whitney U-test across trials in 10/10 rats; and $P = 0.86$, Mann-Whitney U-test across rats). We observed a small but systematic effect of the outcome of the previous trial (correct/error) and the WT of the previous trial (short/long) on absolute WT. Rats tended to wait longer for reward following trials with correct outcome as well as after trials with long WT. These effects did not reach significance when averaging across rats (**Figure 1.S3A**, $P > 0.1$, Mann-Whitney U-test across rats), but were significant in many individual rats (**Figure 1.S3C, D**, $P < 0.05$, Mann-Whitney U-test across trials in 7/10 rats for the effect of previous outcome, and 5/10 rats for the effect of previous WT). These patterns of modulation would be expected if the distribution of temporal reward expectancies, $P_{rew}(t)$, was updated based on the reinforcement history. At the same time, these effects did not lead to significant changes to the C-dependence of WT (**Figure 1.S3E-P**, $P > 0.1$, ANOVA across rats, See Experimental Procedures I).

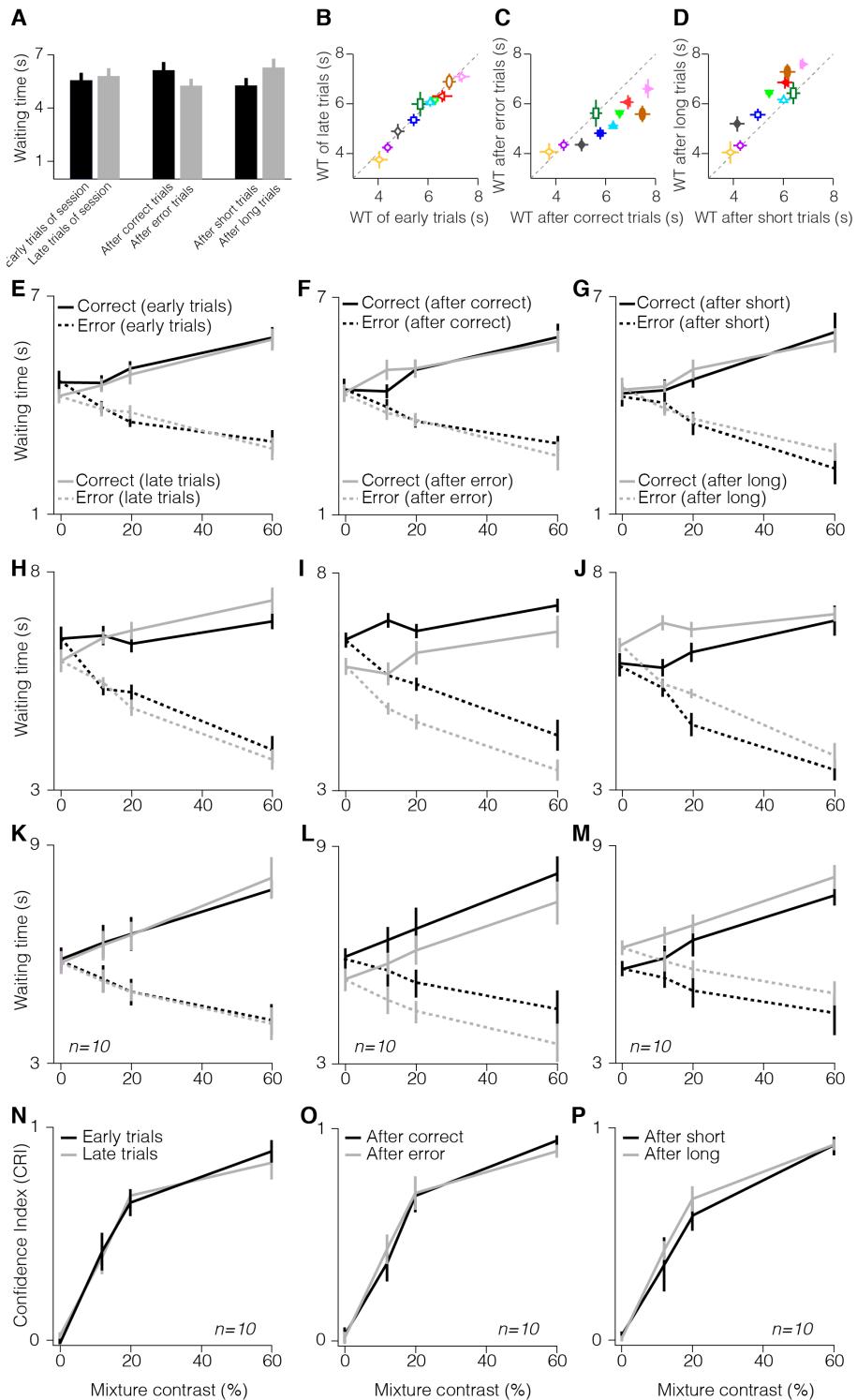


Figure 1.S3 - Dependence of confidence reporting measures on trial history.

(A) Mean waiting times, averaged across rats ($n=10$), as a function of different parameters related to trial history. Left bars: WTs of early trials (initial 30% of trials per session) and late trials (last 30% of trials per session) of session; middle bars: WTs after correct and after error trials; right bars: WTs after trials with short WT (shorter than median waiting time of session) and after trials with long WT (longer than median waiting time of session). For all panels of the figure, error bars are \pm s.e.m across trials or across rats.

(B) Scatter plot of mean waiting time of individual rats for early trials (initial 30% of trials per session) and late trials (last 30% of trials per session) of session. Each data point corresponds to a single rat. Here and for all scatter plots of this figure filled/empty markers indicate

statistically significant/non-significant differences of waiting time between the two different conditions indicated on x and y axes of plots (Mann-Whitney U-test, $P < 0.05$).

(C) Scatter plot of mean waiting time of individual rats when the previous trial had correct or error outcome.

(D) Scatter plot of mean waiting time of individual rats when the previous trial had short WT (shorter than median waiting time of session) or long WT (longer than median waiting time of session).

(E) Mean waiting time of an example rat as a function of odor mixture contrast and trial outcome for early trials and late trials of session. This parameter did not affect the waiting time of this rat (shown with empty purple circle in (B)).

(F) Mean waiting time of an example rat (same rat as in (E)) as a function of odor mixture contrast and trial outcome after correct and error trials. This parameter did not change the waiting time of this animal (shown with empty purple circle in (C)). **(G)** Mean waiting time of an example rat (same rat as in (E, F)) as a function of odor mixture contrast and trial outcome after trials with short and long waiting time. This parameter did not affect the waiting time of this animal (shown with empty purple circle in (D)).

(H, I, J) As in (E, F, G) for an individual rat which its waiting time was not affected by early/late trials but was affected by the previous outcome (correct/error) and previous waiting time (short/long). This rat was indicated by red triangles in (B, C, D).

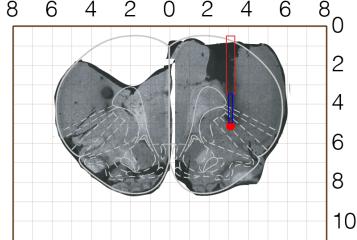
(K, L, M) As in (E, F, G) averaged across all ten rats.

(N, O, P) Confidence-reporting index (CRI) as a function of odor mixture contrast conditioned on the same contingencies as in (B, C, D). CRI was measured by constructing receiver operating characteristics (ROC) curve from the distribution of waiting times of each rat.

Inactivation of orbitofrontal cortex impairs confidence-based waiting times but not choice accuracy

Next, we pharmacologically inactivated OFC and examined decision performance and C-dependent waiting times. Rats were first trained in the task described above. After reaching criterion performance levels we implanted dual cannulae bilaterally in lateral and ventrolateral parts of OFC (**Figure 1.4A** and **Figure 1.S4**, see Experimental Procedures I).

A
Posterior cannulae positioned in OFC
(AP = 2.7 mm)



B
Anterior cannulae positioned in OFC
(AP = 3.2 mm)

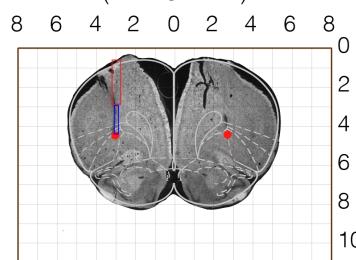
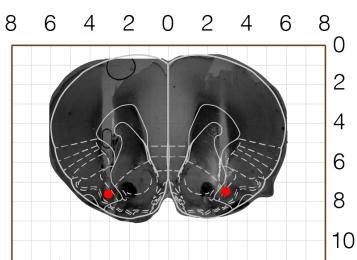


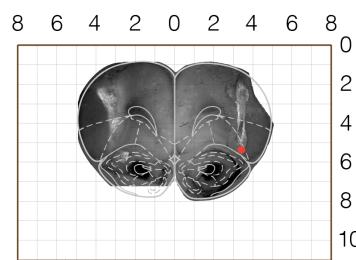
Figure 1.S4-Histology slides

Examples of histology sections showing the position of implanted cannulae in different rats. In 4 rats the implanted cannulae correctly targeted the OFC whereas in other animals, implanted cannulae were located outside the OFC.

C
Cannulae positioned in piriform
(AP = 2.2 mm)



D
Cannulae positioned lateral to OFC
(AP = 3.7 mm)



(A,B) Example slides in which the cannulae correctly targeted the OFC.

(C) Example slide in which cannulae tip positioned too deep, impinging on the pirifrom cortex.

(D) Example slide in which cannulae tip positioned outside (very lateral) the OFC.

Following recovery, on alternate testing days rats ($n=4$) received intra-OFC infusion of either the GABA-A agonist muscimol for silencing neural activity or a saline solution or no injection. Because we found no differences in accuracy, reaction time and WT between saline and no injection sessions ($P > 0.1$, Mann-Whitney U-test across trials in 4/4 rats; and $P > 0.1$ Mann-Whitney U-test across rats), we combined these as the control condition. We found that OFC inactivation did not change sensory discrimination performance (Figure 1.4B and 1.S5A), odor sampling duration or movement time (Figure 1.S5D-I), establishing that it is not required for perceptual decisions (accuracy: $P > 0.1$, bootstrap test on the slope of the psychometric functions in 4/4 rats; and $P > 0.6$, ANOVA across rats; reaction time: $P > 0.1$, ANOVA across trials in 2/4 rats ($P = 0.01$ in other 2 rats); and $P > 0.2$, ANOVA across rats, see Experimental Procedures I).

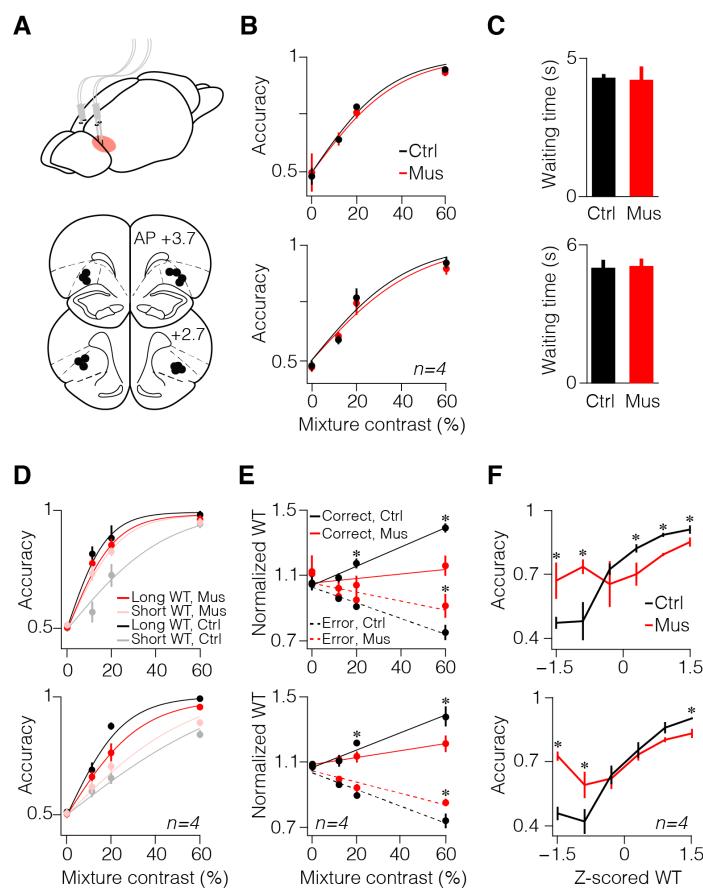


Figure 1.4 - OFC inactivation disrupts confidence-dependent waiting time but not decision accuracy.

(A) Schematic for cannulae implants and anatomical locations of confirmed inactivation sites across rats. See Figure S4 for examples of histology sections.

(B) Decision accuracy as a function of odor mixture contrast for control (saline and no injection combined) and muscimol conditions for the example rat (top) and averaged across rats (bottom). Lines are logistic fits to the data (see Experimental Procedures I). In all panels error bars are \pm s.e.m. across trials or across rats. Cannulae implantation itself had no effect on the decision accuracy (Figure S6).

(C) Mean waiting times for control and muscimol conditions for the example rat (top) and averaged across rats (bottom).

(D) Psychometric functions separated based on WT in the control and muscimol conditions for the example rat (top) and averaged across rats (bottom). Black and gray dots represent long WT (above 70 percentile) and short WT (shorter than 70

percentile) control trials, respectively. Red and pink dots represent long WT (above 70 percentile) and short WT (shorter than 70 percentile) muscimol trials, respectively. Lines represent logistic fit on the accuracy data (see Experimental Procedures I).

(E) Mean normalized WT plotted as a function of odor mixture contrast and trial outcome for control and muscimol conditions for the example rat (top) and averaged across rats (bottom). In order to combine WTs across different sessions of each rat and across rat, normalized WTs were used. For this normalization, the WT in each trial was divided by mean WT of all trials of the session (see Experimental Procedures I). Lines are linear fit to the data. Asterisks indicate significant differences ($P < 0.05$) between individual data points. Cannulae implantation itself had no effect on the WT pattern (Figure 1S6). See Figure 6 for effect of muscimol on WT patterns in rats with cannulae positioned outside OFC. **(F)** Decision accuracy as a function of z-scored waiting time (see Experimental Procedures I) for control and muscimol condition for the example rat (top) and averaged across rats (bottom).

Moreover, average WT was not affected by OFC inactivation (**Figure 1.4C and 1.S5C**; $P > 0.2$, Mann-Whitney U-test across trials in 3/4 rats ($P = 0.01$ in the fourth rat); and $P > 0.8$, Mann-Whitney U-test across rats). However, while psychometric functions of the short and long WT trials had significantly different slopes in the control condition, this difference was negligible in the inactivation condition (**Figure 1.4D**; $P < 0.05$ in 4/4 rats; and $P < 0.01$ across rats, bootstrap test on the slope differences). Moreover, we found that the dependence of WT on stimulus difficulty and outcome was significantly reduced (**Figure 1.4E**; $P < 0.01$, bootstrap test on the slope of the fitted lines in 4/4 rats; and $P < 0.01$, ANOVA across rats, see Experimental Procedures I) without a concomitant change in the mean WT. In addition, accuracy as a function of WT flattened (**Figure 1.4F**; $P < 0.05$, Mann-Whitney U-test for selected time bins across trials in 3/4 rats; and $P < 0.05$; Mann-Whitney U-test for selected time bins across rats) establishing that WT became a worse predictor of performance.

The previous analyses only considered the mean WT patterns and not their variance and distribution. Therefore we next evaluated how well a subject's waiting time report conformed to its actual decision accuracy using type-II receiver operating characteristic (ROC) analysis (Fleming and Dolan, 2010; Kepcs et al., 2008; Rounis et al., 2010), (**Figure 1.5A,B**, see Experimental Procedures I). This confidence-reporting index (CRI) systematically varied as a function of stimulus difficulty (**Figure 1.5C,D**; $P < 0.01$, ANOVA across trials in 4/4 rats) as expected and was significantly reduced by OFC inactivation (**Figure 1.5C,D** and **1.S5B**; $P < 0.01$, ANOVA across trials in 4/4 rats; and $P < 0.05$, ANOVA across rats).

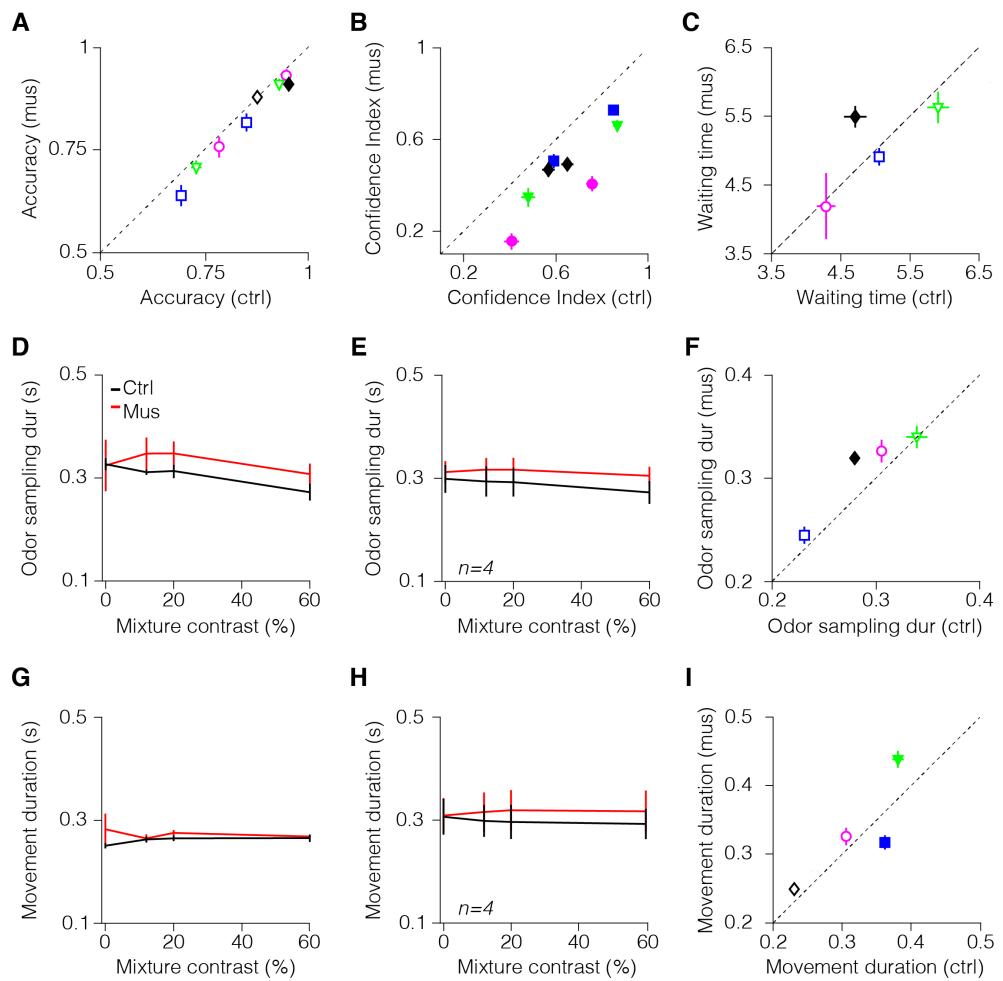
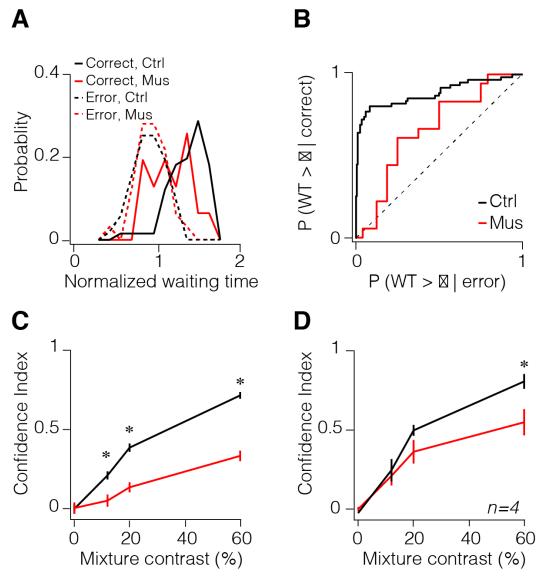


Figure 1.S5 - OFC inactivation disrupts decision confidence but not perceptual decisions in individual rats.

(A) Scatter plot of behavioral accuracy for individual rats for muscimol and control conditions. Each data point corresponds to a single rat for single odor mixture contrast condition (either 20% or 60%). Here and for all scatter plots filled/empty markers indicate statistically significant/non-significant differences for the measured variable between the muscimol and control conditions ($P < 0.05$). For all panels of the figure, unless stated otherwise, error bars are \pm s.e.m across trials or across rats. **(B)** Scatter plot of confidence-reporting index (see Experimental procedure I and Figure 1.5 for detailed definition) for individual rats for muscimol and control conditions. Each data point corresponds to a single odor mixture contrast condition (either 20% or 60%) of a single rat. Error bars are bootstrapped estimates. **(C)** Scatter plot of mean waiting time of individual rats for muscimol and control conditions. Each data point corresponds to a single rat. **(D)** Mean odor sampling duration (defined as the interval between odor valve opening time and the time point on which rat leaves the odor port) for the example rat as a function of odor mixture contrast for muscimol and control conditions. **(E)** Mean odor sampling duration as a function of odor mixture contrast for muscimol and control conditions averaged across rats. **(F)** Scatter plot of odor sampling duration of individual rats for muscimol and control conditions. **(G)** Mean movement duration (defined as the interval between the time of leaving the odor port and the time of entry into a choice port) as a function of odor mixture contrast for muscimol and control conditions for the same rat as in (D). **(H)** Mean movement duration as a function of odor mixture contrast in muscimol and control conditions averaged across rats. **(I)** Scatter plot of movement duration of individual rats for muscimol and control conditions.



odor mixture contrast for control and muscimol conditions \pm s.e.m. across rats.

Figure 1.5 - OFC inactivation reduces the accuracy of confidence report.

(A) Probability distribution of normalized WTs for error and correct (reward omission) trials in control and muscimol conditions shown for 60% odor mixture contrast for the example rat. **(B)** Receiver operating characteristics (ROC) curve computed from probability distributions in (A), as threshold, θ , varied. A rescaled value for the area under this ROC curve is used as the confidence-reporting index (CRI, see Experimental Procedures). **(C)** CRI as a function of odor mixture contrast for control and muscimol conditions for the example rat. Error bars are bootstrapped estimates. **(D)** CRI as a function of

Finally, we considered the specificity of these results to the ventrolateral portion of OFC (vlOFC). We initially excluded 5 rats from our previous analyses where histological examination showed that some of the four cannulae were positioned either too lateral to vlOFC or too ventral reaching the piriform cortex (**Figure 1.S4**). In order to quantify the relationship between the position of cannulae and the behavioral effects, we measured the position of the cannulae relative to the centers of the vlOFC and the piriform cortex (see Experimental Procedures I for details). We then examined confidence reports and perceptual accuracy as a function of the average distance of cannulae relative to the OFC and piriform cortex (**Figure 1.6A**). The perceptual accuracy of rats with cannulae close to the piriform cortex was attenuated by muscimol inactivation, suggesting an important role for the piriform region in our odor-guided decision task (**Figure 1.6B**; $P < 0.05$ in 2/2 rats, bootstrap test on the slope differences). On the other hand, when cannulae were positioned very laterally, outside vlOFC, then we did not observe any effects of inactivation on either perceptual accuracy or the WT pattern (**Figure 1.6C**; $P > 0.2$ in 3/3 rats, bootstrap test on the slope differences). These results specifically implicate the ventrolateral subregion of OFC in confidence reporting.

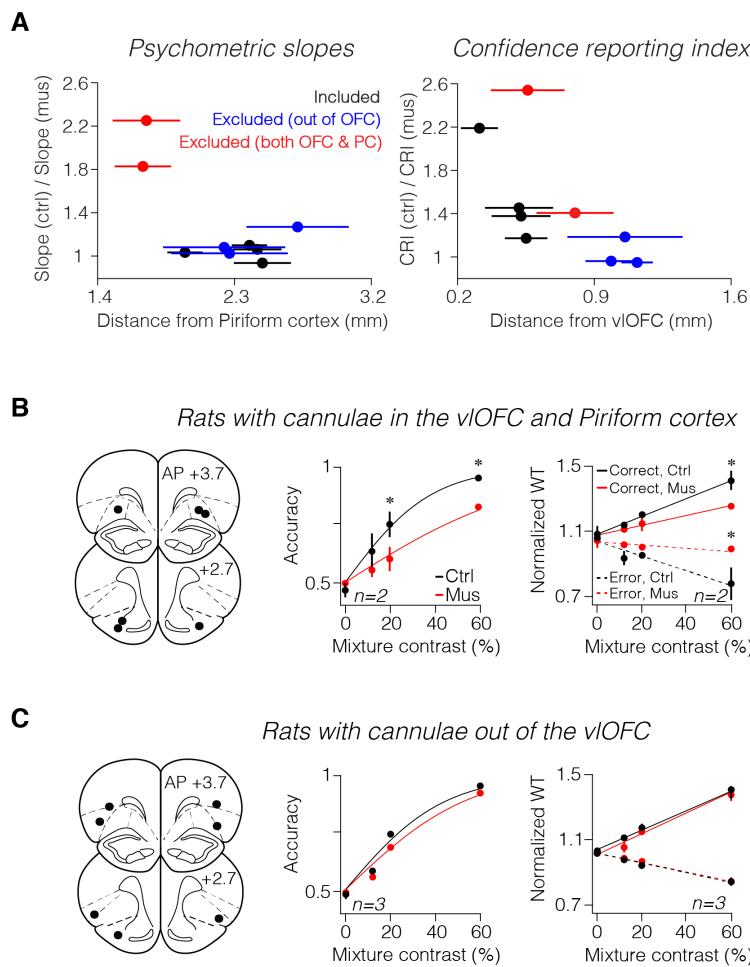


Figure 1.6 - Observed behavioral effects are specific to ventrolateral OFC inactivation.

(A) Behavioral effects as a function of cannula location. (left) Psychometric slope ratios for each rat as a function of cannulae distance from the piriform cortex. (right) Confidence-reporting index (CRI) ratios for each rat as a function of cannula distance from the vOFC. Each dot indicates the average distance measured for each rat (averaged across all visible cannulae tracks). Error bars are s.e.m. across measurements.

(B) Rats with cannulae both in the OFC and piriform cortex ($n=2$). (left) schematic of cannulae positions. (Middle) rats' psychometric functions. (right) rats' WT pattern. Asterisks indicate significant differences ($P < 0.05$) between individual data points.

(C) The same as (B) for rats with cannulae out of the OFC ($n=3$). Compare behavioral effects shown in (B, C) with effects shown in Figure 4B and E.

DISCUSSION

Confidence judgments are usually studied using explicit self-reports in humans and are taken at face value. In order to study non-human animals a different approach is required (Kepecs and Mainen, 2012). We introduced a new post-decision gambling task that makes confidence reports valuable for animals and allowing experimenters to collect choices and confidence reports from the same trials (Kepecs and Mainen, 2012; Middlebrooks and Sommer, 2012). This is an advantage compared to opt-out tasks in which animals are presented with a third choice that provides a guaranteed but smaller reward. Opt-out choices may be made in epochs when the attentional or motivational state of an animal is reduced, so that if an animal is monitoring these state changes, it could prefer to opt out of the perceptual decision. For these reasons,

opt-out designs are not ideal for studying decision confidence because in these tasks each trial only provides either a perceptual or an opt-out choice, making it difficult to rule out behavioral mechanisms that do not require uncertainty monitoring. The time investment gambling task described here had fundamental similarity to the restart task we previously employed in which rats could abort the current trial to restart a new trial (Kepcs et al., 2008). However, the restart task provided only a binary measure of decision confidence (i.e. stay or restart). Consequently, another feature of current task design is that waiting times (WTs) served as continuous wagers (instead of binary bets). This is preferable in order to mitigate the problem of finding the optimal payoff matrix for binary bets that depends on animals' internal costs and valuations (Clifford et al., 2008; Middlebrooks and Sommer, 2011; Schurger and Sher, 2008).

In order to establish that WTs could serve as indices of confidence, we compared rats' WT patterns to a normative model of decision confidence. First, we showed that the optimal time to wait depends monotonically on the initial reward probability for each trial (**Figure 1.2**). In perceptual decisions, reward probability can be estimated based on the confidence associated with a decision. Second, we derived three predictions for decision confidence and compared these to WTs. As expected for a proxy of confidence, we found that WTs (i) correlated with the slope of psychometric functions, (ii) predicted decision accuracy and (iii) showed a characteristic dependence on signal-to-noise ratio and outcome (**Figure 1.3**). This allowed us to interpret our findings in the context of a normative model rather than a semantic definition of confidence. From a computational standpoint, the observed WTs could only be explained by models in which the variable $P(\text{correct}|\text{evidence})$, i.e. confidence, is taken into consideration. Rats' WTs are determined not only by decision confidence but also by estimated reward delivery time and other reinforcement-related factors (**Figure 1.S3**). Nevertheless, the accurate computation of such reward expectation is only possible by incorporating confidence information.

Inactivation of the ventrolateral portion of OFC disrupted the confidence-dependence of WTs without a change in decision accuracy or mean waiting time (**Figure 1.4-1.6**). These results provide evidence that an intact OFC is necessary for reporting confidence but not for perceptual decision-making under uncertainty. Beyond establishing an anatomical locus for confidence judgments, the results also show that confidence reporting and the computation of perceptual decisions are at least in part distinct processes localized to different brain regions. From this perspective, our findings reinforce recent observations regarding the role of pulvinar in the representation of perceptual confidence (Komura et al., 2013). Using an opt-out task, these authors showed that inactivation of pulvinar increases monkeys' opt-out choices in the wagering task without affecting perceptual categorization. However, for reasons discussed above, this experimental design could not rule out alternative possibilities. For instance, pulvinar inactivation could cause either lower risk-taking propensity or reduced attention, both leading to an increased opt-out behavior (Kepecs, 2013).

Our findings leave open the question of whether OFC locally computes confidence or instead receives confidence signals from other areas (Insabato et al., 2010a, 2010b; Kiani and Shadlen, 2009; Komura et al., 2013) but we consider it likely that choice and confidence are computed together and represented in regions important for perceptual decision-making and then relayed to OFC. Centralization of confidence signals in OFC may be useful in order to have a central region to monitor confidence levels, alongside other reward-related variables, regardless of perceptual modality. Neuronal signals related to metacognitive monitoring have been observed in several subregions of the frontal cortex (Fleming and Dolan, 2010; Kepecs et al., 2008; Lau and Passingham, 2006; De Martino et al., 2013; Middlebrooks and Sommer, 2012; Persaud et al., 2007; Rolls et al., 2010a, 2010b; So and Stuphorn, 2012; Tsujimoto et al., 2010; Yokoyama et al., 2010), as well as in parietal cortex (Kiani and Shadlen, 2009) and thalamic nuclei such as pulvinar (Komura et al., 2013) suggesting that metacognitive representations may be widespread in the brain. Our results suggest that

OFC may integrate distinct sources of information, and similar to its role in value-based decisions, may provide outcome predictions based on confidence monitoring processes.

Previous findings have implicated OFC in representing reward expectations (Rolls and Grabenhorst, 2008; Schoenbaum and Roesch, 2005; Wallis, 2007) and in goal-directed behavior across species (Morrison et al., 2011; Rolls and Grabenhorst, 2008; Schoenbaum et al., 2009; Wallis, 2007). Because decision confidence is also critical for computing the value of the current decision outcome (Hare et al., 2008), our results are consistent with a role for OFC in outcome valuation. OFC lesions are also known to impair the devaluation of reward outcomes, reversal learning and increase impulsivity (Bechara et al., 1997; Berlin et al., 2004; Burke et al., 2009; Mar et al., 2011; Noonan et al., 2010; Rolls and Grabenhorst, 2008; Rudebeck and Murray, 2008; Schoenbaum et al., 2002, 2009; Wallis, 2007; Walton et al., 2010), all of which reflect a compromise in the relative potency of explicitly imagined outcomes, as opposed to routine habits, in driving decisions (Balleine, 2011). These are consistent with our observations that OFC inactivation only affected WT wagering behavior and not well-learned decisions. Inactivation of OFC might have impaired general reward expectations and the motivation to wait for the reward (Mar et al., 2011; Noonan et al., 2010). However, WT after OFC inactivation was only reduced for correct trials, and increased for incorrect trials (**Figure 1.4**) suggesting that a disruption of reward expectation could not by itself account for the data. The fact that OFC inactivation did not affect the mean WT suggests that the animals' ability to estimate the elapsed time remained intact. Moreover, rats' movement times were not different between the muscimol and control sessions (**Figure 1.S5**) and were not different when comparing pre and post cannulae implantation (**Figure 1.S6**) implying that the observed behavioral patterns could not be attributed to the inactivation nor to the lesion of motor-related structures along the cannulae walls.

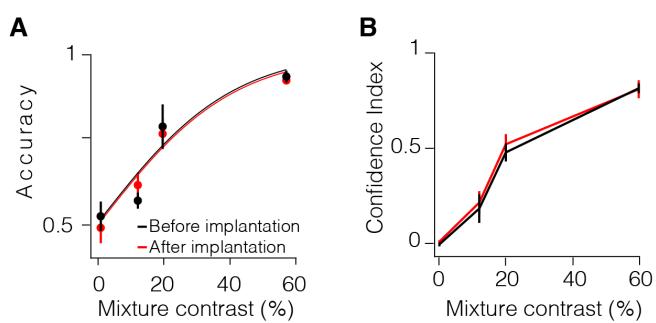


Figure 1.S6 - Cannulae implantation does not affect behavioral performance.
(A) Averaged choice accuracy and psychometric functions of rats before and after the Cannulae implantation. **(B)** Confidence reporting index (CRI, See Experimental procedure and Figure 5 for definition of CRI) before and after the Cannulae implantation.

The observation that following OFC inactivation rats failed to adjust their waiting time based on the decision is consistent with the broader notion that an intact OFC is necessary for some aspects of reward-maximizing choice behavior (Padoa-Schioppa, 2011; Wallis, 2007). Confidence is also a form of uncertainty, therefore, our results are broadly consistent with observations demonstrating that OFC is involved in representing uncertainty and risk in humans (Critchley et al., 2001b; Fleming and Dolan, 2010; Hsu et al., 2005; De Martino et al., 2013; Rolls et al., 2010a; Tobler et al., 2007), monkeys (O'Neill et al., 2010) and rats (Kepecs et al., 2008; Roitman and Roitman, 2010).

In summary, our results support the view that OFC is particularly important for reward-based behaviors when values are inferred, for instance using model-based reinforcement learning algorithms (Daw and Doya, 2006; Hampton et al., 2006; Jones et al., 2012; McDannald et al., 2011; Wilson et al., 2014), rather than when values are stored based on previous experiences. This is because, the estimation of decision confidence is an example of the computation of an inferred value based on a hidden belief state. Consequently, the role of OFC in confidence monitoring can be viewed as a second-order and metacognitive process that fits into the broader conception that OFC is critical for making on the fly predictions about behaviorally important outcomes.

EXPERIMENTAL PROCEDURES - I

Subjects

A total of 10 male Long-Evans rats were used for the experiments. Data from all rats were used for the quantification of confidence reporting behavior. 9 rats underwent the cannulae implantation surgery and based on anatomical localizations of implanted cannulae (**Figure 1.6** and **1.S4**), data collected from 4 rats were used for investigating the effect of OFC inactivation on the confidence reporting behavior.

Rats were motivated by water restriction and had unlimited access to food. All procedures involving animals were carried out in accordance with National Institutes of Health standards and were approved by the Cold Spring Harbor Laboratory Institutional Animal Care and Use Committee.

Behavior

Behavioral task and training

The apparatus has been described previously (Kepes et al., 2008; Uchida and Mainen, 2003). Rats self-initiated each experimental trial by introducing their snout into the central port where odor was delivered. After a variable delay, drawn from a uniform random distribution of 0.2–0.5 s, a binary mixture of two pure odorants, S(+)-2-octanol and R(-)-2-octanol , was delivered at one of 7 concentration ratios (80/20, 60/40, 57/43, 50/50, 43/57, 40/60, 20/80 creating odor mixture contrast of 0, 12, 20 and 60 %) in pseudorandom order within a session. After a variable odor sampling time up to 0.7 s, rats responded by withdrawing from the central port, which terminated the delivery of odor, and moved to the left or right choice port. Choices were rewarded according to the dominant component of the mixture, that is, at the left port for mixtures A/B > 50/50 and at the right port for A/B < 50/50. For trials with 50/50 odor mixture, the reward was randomly assigned to

one of the choice ports. A variable reward delay period after entry into the choice port was introduced. For correct choices, reward was delivered between at least 0.5 s after entry into the choice port and up to 8 s. The reward delay was drawn from an exponential distribution with decay constant equal to 1.5 (**Figure 1.1D**) resulting in a relatively constant level of reward expectancy over a range of delays (i.e. flat hazard rate). In a small fraction of correct choice trials distributed pseudorandomly throughout the behavioral session (10% - 15% of correct trials) rewards were omitted. These reward omission trials were distributed so as to never occur on the consecutive trials. As the rat spends time to consume the water in the rewarded correct trials, we used the reward omission trials to measure WT in the correct trials.

In order to perform the task described above, rats went through a multistep training procedure typically lasting 6-8 weeks starting with imperative trials, moving to choice trials and gradually introducing choice trials with low odor mixture contrast.

Surgery and inactivation procedures

Surgery

All surgical procedures were carried out under aseptic conditions. Anesthesia was initiated with inhalation of 2.5% isoflurane (Vetland, Louisville, KY) and retained with intraperitoneal injections of ketamine (50 mg/kg) and medetomidine (0.4 mg/kg). After craniotomy, dual guide cannulae (26-gauge Plastics One, Roanoke, VA) were stereotactically implanted in each hemisphere targeted 1.5 mm above OFC (AP+3.2, ML+/-3.2, DV+2.8 from dura and AP+4.1, ML+/-2.8, DV+1.8 from dura). Dual stainless steel stylets were inserted into the guide cannulae to ensure patency (protruding 0.5 mm below the tip of the guide cannulae).

Pharmacological inactivation

Temporary inactivation was achieved via localized injections of γ -aminobutyric acid (GABA_A) receptor agonist muscimol (Sigma Alderich)

under light anesthesia induced by 2% isoflurane (for about 6 min during which hind leg reflex never disappeared over the course of infusion). On each testing day the stylets were replaced with dual injector cannulae (33-gauge, Plastics One) protruding 1.5 mm below the tip of guide cannulae. One minute after proper placement of the injectors, muscimol (0.05 µg in 0.4 µl) or sterile saline (0.9%; 0.4 µl) was injected over a 5-minute period. Fluid was infused via 0.38 mm diameter polyethylene tubing (Intramedic, New York, NY) attached to the injector on one end and to a 2 µl Hamilton syringe (Hamilton, Reno, NV) on the other end. The syringe was driven with a syringe pump (Harvard Apparatus, MA).

Injections were monitored by observing the movement of a small air bubble in the tubing to confirm that fluid was moving.. After infusions were complete, the injector cannulae were left in place for 2 minutes and then replaced with stylets to maintain cannulae patency. Behavioral testing began about 30 minutes after infusion. It has been shown that the maximal extent of muscimol spread, using this procedure, was 1.5 to 2 mm within 10-20 minutes of injection (Martin and Ghez, 1999).

Histology

Once experiments were complete, rats were deeply anesthetized and then transcardially perfused with 4% paraformaldehyde. Brains were removed, postfixed, and coronal sections of 50 µm were made using a fixed-tissue vibratome (VT1000S, Leica Instruments, Germany). Only animals in which at least 3 of the 4 cannulae were located within the lateral and ventrolateral portions of OFC were included in our analysis. We determined that 4 out of 9 implanted animals had correct cannulae positions while the others extended either ventrally into the piriform cortex or caudally into the striatum (Figure 6 and S4).

Computational model

Confidence estimation model

To predict the expected patterns of decision confidence we used a simple model for two-alternative decisions. We used a signal detection theory

framework where each choice and its associated confidence could be estimated by comparing the sampled stimulus and the decision boundary. The boundary was fixed at 50%. In this framework, the choice in each trial is computed by comparing stimulus and boundary. We modeled the stimulus as the percentage of one of the components in the mixture henceforth denoted by m . The internal representation of the stimulus m' is different from the actual value m :

$$m' = m + \xi \quad (8)$$

Here ξ is the Gaussian variable with zero mean and the standard deviation of σ : $p(\xi) = e^{-\xi^2/2\sigma^2} / \sqrt{2\pi\sigma^2}$. The origin of noise ξ is two-fold: it may be contributed by the external uncertainties in the stimulus as well as the internal sources of error. From fits to experimental data we obtained an estimate of $\sigma \approx 18\%$, i.e. internal and external sources of noise in their strength are equivalent to about 18% of the fraction of one components in the mixture. In each trial, it is assumed that the value of the internal representation of the stimulus determines the response of the observer. The values of m' exceeding the decision boundary $b=50\%$ result in response to the right, while the values $m' < 50\%$ produce a left response. For a given external stimulus m , the fraction of right responses i.e. the psychometric function is given by

$$P(m) = \frac{1}{2} \left[1 + \operatorname{erf} \left(-\frac{m-b}{\sigma\sqrt{2}} \right) \right] \quad (9)$$

In each trial, the distance between the internal representation of stimulus m' and boundary provides an estimate of decision confidence, C . Decision confidence in our approach is defined as the probability of making the correct decision $C = C(m')$. Decision confidence however is a function of the internal representation of the stimulus m' which is different from the external value (Equation 8). It is not difficult to see

that for a simple decision task, the probability of being correct could be estimated using Equation 7.

It is important that, in each individual trial, the same internal representation of the stimulus m' that determines response (response to the right occurs when $m' \geq 50\%$) is used to evaluate the decision confidence through Equation 7. The internal representation of the stimulus m' varies from trial to trial even if the stimulus mixture m is fixed. Response accuracy therefore becomes coupled with decision confidence. Because $C = C(m')$ is an internal variable, it is not available for direct measurement. Instead, it could be assessed through the time spent by the observer in the reward port. In the computational model, we used Equation 6 describing an ideal observer. Note that Equation 6 predicts that there are only two conditions when WT goes to infinity (i.e. waiting never terminates): when the opportunity cost, κ , is zero (waiting has no cost) or when a rat is completely confident about its choice $C = 1$. However, there is always some reward to be gained in future trials and because of reward omission trials a rat cannot be completely certain of reward, hence WT is always finite.

Fitting the model to behavioral data

To fit our model to rats' behavioral data, we estimated two parameters. First, the width of the total noisy distribution (σ , made up of sensory and internal noise) used to calculate model's trial by trial choice and confidence. Second, opportunity cost (κ) which, alongside confidence and reward delay distribution (Equation 6) is used to calculate model's WT. Parameter estimation was done using a maximum likelihood method, implemented using MATLAB's *fminsearch* function. To avoid local minima, we re-ran *fminsearch* 1000 times using random starting parameter values and selected the set of parameter estimates with the smallest mean squared error.

Starting from rat's psychometric curve, we estimated one parameter (σ) that minimized the mean square error of choice predictions. We then used the estimated σ to calculate the intermediate variable, decision

confidence (C), in each trial (**Figure 1.3A**, middle panel) and subsequently estimated one other parameter, κ (Equation 6) which could minimize the difference between rat's and the model's WT distribution (Figure 3A, right panel). Although this model fit the mean WTs per condition well, in order to fit to the entire WT distribution we also assumed ‘scalar timing’. In other words, we assumed that a rat's estimation of elapsed time carries uncertainty and in particular that the standard deviation scales with elapsed time (Gibbon, 1977; Gibbon et al., 1997; Janssen and Shadlen, 2005). This implies that a time t is perceived at time $t \pm \sigma(t)$, where

$$\sigma(t) = \phi \cdot t \quad (10)$$

where ϕ is the coefficient of variation or Weber fraction (Gibbon, 1977). Consistent with previous findings, we set $\phi = 0.3$ (Gibbon et al., 1997; Janssen and Shadlen, 2005). Therefore, for the fitting the model's WT distribution was blurred with a normal distribution whose standard deviation was proportional to the elapsed time (**Figure 1.3A**, right panel). Lines in **Figure 1.3B-D** show predictions of the model with parameters optimized to fit rat's accuracy curve and WT distribution.

Analysis of behavioral data

We collected 68243 trials from 10 rats as following: 24032 control trials (Saline injection or no injection) and 11106 muscimol trials from 4 rats (79 sessions, in average 445 trials per session per rat, min trial per session=295, max trial per session=654, min session per rat = 15, max session per rat = 29). These data were included in analysis of confidence-related WTs as well as the muscimol experiment. 23473 control trials and 9632 muscimol trials from 6 rats (5 of them implanted, 81 sessions, in average 408 trials per session per rat, min trial per session=281, max trial per session=701, min session per rat = 12, max session per rat = 18). Due to incorrect position of cannulae revealed by histological examination (**Figure 1.S4**), these data were only used in the analysis of confidence-related WTs and in Figure 6.

In general, we used nonparametric Mann-Whitney U-test for single comparisons and one or two-way ANOVA, post-hoc test adjusted, for multiple comparisons. Bootstrap test used for the comparing fit parameters in **Figure 1.3E, 1.4B,D,E** and for comparisons shown in **Figure 1.5C** and **1.S5B**. For statistical analysis across rats, averaged data for each rat was used. However, the large number of trials collected for each rat also enabled us to examine the significance of behavioral effects for each subject separately. For such analyses on single rats, statistical tests were performed across all trials collected for each animal. Asterisks in figures illustrate statistically significant ($P < 0.05$) differences for individual data points using Mann-Whitney U-test (bootstrap test was used for **Figure 1.5C**). Filled/empty markers in scatter plots indicate significant/non-significant differences tested using Mann-Whitney U-test (bootstrap test was used for Figure **1.S5B**). Unless stated otherwise, error bars in figures indicate standard error of mean (s.e.m) across trials for individual animals or across rats for the population data.

Perceptual accuracy and reaction time data

For illustration proposes only, we fit behavioral choice data (probability of choosing left port) as a function of odor concentration (%A) to a logistic function of the following form (**Figure 1.1B**):

$$Accuracy = \frac{1}{1 + e^{-(\alpha + \beta \times Odor\ mixture)}} \quad (11)$$

where α is a measure of choice bias and β reflects perceptual sensitivity.

We fit behavioral accuracy data as a function of odor mixture contrast to a logistic function of the following form (**Figure 1.3E** and **1.4B, D**):

$$Accuracy = \frac{1}{1 + e^{-(\beta \times Odor\ Contrast)}} \quad (12)$$

where β reflects perceptual sensitivity (i.e. psychometric slope), with higher values implying increased sensitivity.

Waiting time data

Waiting time data exhibited small variation across sessions and subjects. Therefore, for each rat the WTs of each session were normalized to the mean of the WT of that session (normalized WT). Other possible ways to normalize the data (normalization to the median WT of the session or normalization to the mean/median of the WT for odor mixture contrast = 0) resulted in very similar findings. For illustration proposes, non-normalized WT data was used in Figure 3.

For Figure 4F, z-scored WT (Equation 13, below) was used in order to compute the conditioned accuracy graphs (see below). However, using normalized WT (instead of z-scored waiting time) showed comparable results.

$$Z\text{scored } WT_{trial} = \frac{WT_{trial} - \mu(WT_{session})}{\sigma(WT_{session})} \quad (13)$$

We fit WT data as a function of odor mixture contrast and trial outcome to a linear function of the following form (**Figure 1.3G, 1.4E**):

$$\text{Normalized } WT = \alpha + (\pm \beta \times \text{Odor contrast}) \quad (14)$$

where β indicates the slope of change in the normalized WT as a function of the odor mixture contrast and its sign (-/+) indicates error/correct outcomes respectively.

WT-Conditioned accuracy measures

In order to estimate rats' decision accuracy as a function of WT, we assumed that WTs for correctly performed reward omission trials (which were pseudorandomly distributed) were a good representative for the distribution of all correctly performed trials. Therefore the z-scored WT data (Equation 13) were expanded to all correct trials (taking into account the odor stimulus identity) and WT-conditioned accuracy functions were computed (Figure 1.4F).

Confidence-reporting index (CRI)

An objective measure of confidence reporting ability can be computed based on the type II receiver operating characteristic (ROC) curve, which quantifies how well a subject's confidence report conforms to its actual decision accuracy. For each animal for each of the experimental conditions (muscimol vs. control) the probability distribution of the normalized WT for the error trials and correct trials were first computed (**Figure 1.5A**). The receiver-operating curve was then generated for each of the experimental conditions (Figure 5B), which indicates $P(\text{WT} > \theta | \text{correct})$ as a function of $P(\text{WT} > \theta | \text{error})$, where θ refers to the threshold which was varied in order to construct the ROC curve. The confidence-reporting index (CRI) is the rescaled measure of the area under the ROC curve, so that values close to zero indicate poor confidence and values close to 1 indicate perfect decision confidence (**Figure 1.5C,D, 1.6A, 1.S3N-P and 1.S5B**).

Effects of trial history on waiting time and confidence reporting measures

Apart from decision confidence, the waiting time at the choice ports also depends on when an animal is expecting the reward delivery. We were interested to determine the extent to which WT pattern and CRI, our measures of confidence report, were affected by trial history. Figure **1.S3E-G** show an example rat in which none of the mentioned parameters affected its mean WT. Consequently, the WT patterns are overlapping. **Figure 1.S3H-J** show an example rat in which outcome as well as WT of previous trial affected its absolute WT. As a result, while WT patterns show a general shift, the confidence-dependent WT patterns are robust. When averaging across rats, WT as a function of odor mixture contrast and outcome did not vary between the beginning and end of a session, after correct and error trials or after long and short WTs (**Figure 1.S3K-M**; $P > 0.1$, ANOVA across rats). Similarly, confidence-reporting index (CRI) as a function of odor mixture contrast did not vary between the beginning and end of a session, after correct and error trials or after long and short WTs (**Figure 1.S3N-P**; $P > 0.2$, ANOVA across rats).

**NEURAL CORRELATES OF
CONFIDENCE AND CONFIDENCE
REPORT IN ORBITOFRONTAL
CORTEX AND VENTRAL STRIATUM**

SUMMARY

Difficult decisions can occur because stimuli are hard to perceive or because the rules of what should be done given a certain stimulus are uncertain to the decision maker. We would like to understand how this second form of uncertainty is represented by the brain and may be assessed and used for adaptive behavior. Neural correlates of perceptual decision confidence have been previously found in the orbitofrontal cortex (OFC) of rats. OFC and ventral striatum (VS) are two brain regions implicated in behavioral supervision and outcome evaluation. To better understand the role of OFC and VS in the computation of decision confidence and behavior adaptation tuned by confidence signals we have recorded single unit activity from these two regions from rats performing a two-odor categorization task with a post decision time-wager confidence report – waiting time. We have found populations of cells in both OFC and VS whose activity was correlated with decision confidence and waiting time, in different epochs of a trial. These results have further explored the functions of OFC in confidence based guided decisions and added the basal ganglia to the circuitry involved in decision confidence computations.

INTRODUCTION

Humans have the ability to make judgments about their own decisions. Because the outcome of important decisions is sometimes not readily available - think of a marriage engagement or a career plan- the capability of evaluating the quality of decisions and act according to it is of big importance (Harvey, 1997; Johnson and Fowler, 2011) When waiting for a outcome a decision-maker which is not very confident (more uncertain) can opt for giving up waiting and try a fresh start. On the contrary, when confidence is high (uncertainty is low), to persevere in the promise of a reward might be the best thing to do. We are

interested on understanding how does the brain compute decision confidence and generates confidence-driven adaptative behaviors.

Non-human animals are also capable of self-confident judgments (Smith et al., 2003). In monkeys a number of studies, within the scope of visual guided decision making, have found neuronal correlates of decision confidence in cortical areas such as the parietal cortex (Kiani and Shadlen, 2009), the supplementary eye field (Middlebrooks and Sommer, 2012), the pre-motor cortex (Martinez-Garcia et al, 2014) and in the pulvinar nucleus of the thalamus (Komura et al, 2013)

In rodents neural correlates of decision confidence were only found so far during a reward anticipation period in the orbitofrontal cortex (OFC) of rats (Kepecs et al., 2008) which were performing a two-odor categorization task. No confidence report was included in this previous task so it was only suggested that the confidence correlates could have behavioural significance. In a following study (Lak et al., 2014) inactivation of OFC of rats, behaving in a modified version of the categorization task which included a confidence report, disrupted a waiting-based confidence judgment, without affecting the odor categorization. This further revealed a role for the OFC in metacognitive monitorization of perceptual decisions. Nonetheless it is still not well understood how the neural correlates of decision confidence are related to the behavioural report of confidence. Moreover, the dynamics of the confidence signals are still unclear. Is the OFC computing a decision confidence estimate right after the perceptual decision was made? And can the waiting-time based confidence judgment be predicted as early on by OFC neural activity?

OFC representations have been thoroughly implicated in different aspects of outcome expectations (reviewed in Padoa-Schioppa and Cai, 2011; Schoenbaum et al., 2011; Schultz et al., 2011; Wallis, 2012) and decision confidence has been proposed to be one of the variables that animals take into account for generating these expectations (Mainen and Kepecs, 2009). In respect to behavioural adaptation, OFC activity was found to be necessary for a shift from habitual to goal-directed actions

following outcome revaluation (Gremel and Costa, 2013) which involved dynamic interactions between corticostriatal circuits of OFC and (dorsal) striatal regions.

Ventral striatum (VS), specially the nucleus accumbens shell, is a striatal target of OFC projections (Schilman et al., 2008). These two regions, OFC and VS, are thought to work in concert to guide optimal courses of actions that ultimately lead to rewards (Hare et al., 2008). More specifically it was speculated that signals from OFC bias VS activity with a flow of reward-contingency information (Simmons et al., 2007) and that VS incorporates information about features of expected outcome, signaled from OFC (McDannald et al., 2011).

While the study of this corticostriatal circuit in the scope of reinforcement learning is prolific (reviewed in Balleine and O'Doherty, 2010; Ito and Doya, 2011) this is not the case for the framework of perceptual decision confidence. The only data available involving VS and perceptual decision confidence comes from human functional imaging studies where voxels in VS showed activation positively correlated with reports of decision confidence (Daniel and Pollmann, 2012; Hebart et al., 2014).

To better understand the role of OFC and VS in the computation of decision confidence and behavior adaptation tuned by confidence signals we have recorded single unit activity from these two regions from rats performing a two-odor categorization task with a post decision time-wager confidence report – waiting time. We found populations of cells in both OFC and VS whose activity could predict outcome in a difficulty dependent manner -hence correlated with confidence. The activity of these populations also correlated with the behavioural report of confidence. Furthermore these populations were differently constituted along the trial, with different individual neurons predicting outcome or waiting time depending on the trial epoch. OFC outcome predictive cells were more strongly correlated with waiting time just after decision, and VS cells while anticipating reward.

RESULTS

Odor categorization task with confidence report

Six rats were trained to perform a version of a classical two alternative forced choice odor categorization task (Kepecs et al., 2008; Uchida and Mainen, 2003), which includes a post-wagering report of confidence (Lak et al., 2014). In this task (**Figure 2.1A**, See Experimental Procedures II & III) water deprived animals had to insert their snout into a centre odor port where they received a two-odor mixture stimulus after a 0.2-0.5s delay. They were then free to move to one of two lateral choice ports in order to receive a water reward for correct decisions. Correct decisions depended on the majority odor in the mixture and the difficulty of each choice was manipulated by varying the odor contrast of the mixture delivered. Rat's choices and accuracy depended on the difficulty of the stimulus delivered (**Figure 2.S1**). Upon inserting their snout in the correct choice port rats had to wait for a reward to be delivered, which occurred after a delay of 0.5s to 8s, randomly drawn from a negative exponential distribution with decay of 1.5s. After water consumption they could reinitiate a new trial almost immediately. Incorrect choices (error trials) were not rewarded nor signaled and in 10% of the correct choices water was omitted (catch trials). In both error trials and catch trials the time that rats were willing to wait at the choice port (waiting time) was our behavioural proxy for confidence. Waiting time varied as a function of the probability of having done a correct choice given stimulus difficulty – higher waiting times occurred on average for correct easy choices (80A:20B or 20A:80B mixtures), and lower waiting times for error easy choices (**Figure 2.1B-C**). It was also predictive of choice accuracy (**Figure 2.1D-E**), indicating that it was a good trial-by-trial gradative report of decision confidence, as reported previously (Kepecs and Mainen, 2012; Lak et al., 2014).

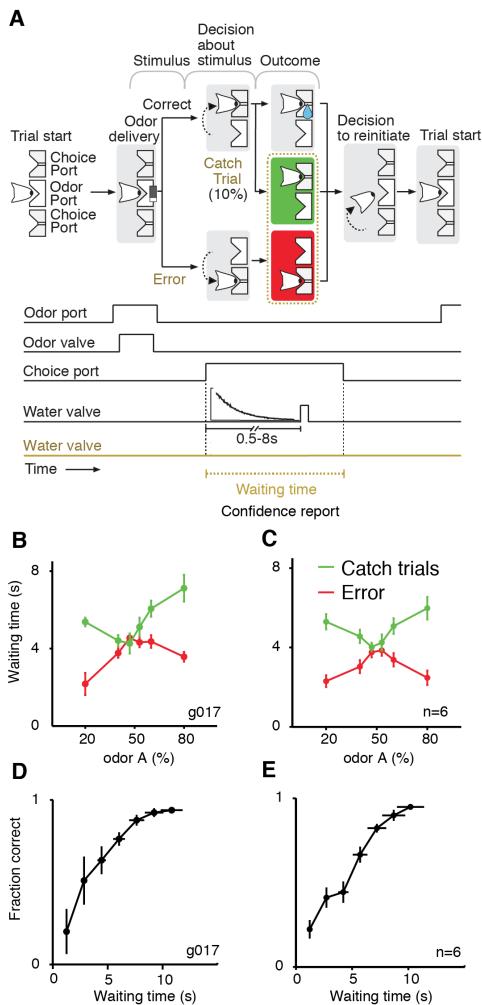


Figure 2.1 – Odor categorization task with confidence report

(A) Schematic of the behavioral paradigm. To start a trial, rats entered the central odor port and after a pseudorandom delay of 0.2–0.5 s a mixture of odors was delivered. Rats moved to one of the lateral choice ports and in correct trials they waited for a drop of water to be delivered after a pseudorandom delay, drawn from a exponential distribution with decay of 1.5s, 0.5s offset and 8s maximum. In error trials no water was delivered, and so was the case for a small percentage of correct trials – catch trials. The time rats were willing to wait with their snout in the reward port, in both error trials and catch trials – waiting time – was considered to be the confidence report for this task. In these trials rats were able to start a new trial after 1s from leaving the reward port.

(B-C) Waiting time as a function of odor-mixture and outcome. Mean waiting time for (B) an example rat or (C) the population of $n=6$ rats, conditional to different delivered odor mixtures, and different outcomes, catch (green) and error (red) trials.

(D-E) Waiting time predicts accuracy. Mean accuracy, as fraction of correct trials conditional on binned waiting times, for (D) example rat or (E) population of $n=6$ rats. Circles and error bars are mean and s.e.m, respectively, across sessions for the same rat in B and D, or across $n=6$ rats, in C and E.

Outcome predictive neurons in ventral striatum and orbitofrontal cortex

Following extensive behavioral training rats were chronically implanted with a tetrode drive, targeting the left hemisphere orbitofrontal cortex (OFC) and the left hemisphere ventral striatum (VS) (see Experimental Procedures). Tetrode tracks were recovered after termination of the experiment using standard histological techniques (see Experimental Procedures) and we considered only data obtained from tetrodes located in the targeted regions during the recording sessions. (**Figure 2.2A-B**). Single cells were isolated offline using manual clustering methods (see

Experimental Procedures), and only units with good isolation and recording stability across the session were further analyzed.

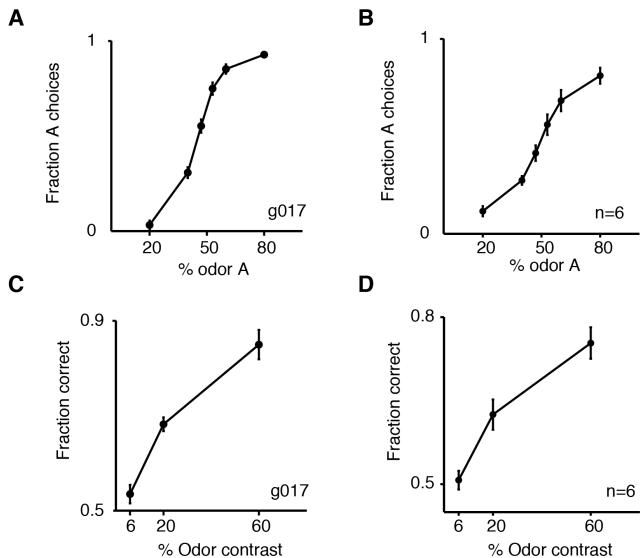


Figure 2.S1 – Behavioral performance in odor categorization task

(A-B) - Performance in the categorization of a mixture of two odors for (A) a example animal and (B) for a population of 6 rats.

(C-D) - Categorization accuracy as a function of odor contrast, for (C) an example rat, and (D) for a population of 6 rats.

Averages and error bars (s.e.m.) were calculated across different behavioral sessions of example rat, or across the population of 6 rats.

As reasoned before (Kepcs et al., 2008) neural activity related to the subject's uncertainty in the outcome of a choice should occur while the subject is anticipating the trial outcome. We also reasoned that neural activity related to our confidence report (the amount of time rats are willing to wait for a reward) should be observed as early as in the beginning of the waiting period. Following this, our first approach was to analyze neural activity during a 500ms window after entry into a choice port, the outcome anticipation period, during which reward delivery is delayed in our task (**Figure 2.2C**).

We have found neurons in the VS, and in the OFC with neural activity that anticipated trial outcome during the outcome anticipation period.

Figures 2.2D and 2.2F depict a VS neuron that fired more intensively during error trials compared with correct trials. This cell is an outcome predictive neuron whose firing rate is also modulated by the difficulty of the categorization, (**Figure 2.2H**). Its firing pattern resembles an “x-pattern” observed previously (Kepcs et al., 2008) which is a signature for an uncertainty coding neuron (Insabato et al., 2010b; Kepcs and

Mainen, 2012). **Figure 2.2E** and **2.2G** depict an OFC neuron which is also an outcome predictive neuron, modulated by the difficulty of categorization, but with a not so clear “x-pattern” (**Figure 2.2I**).

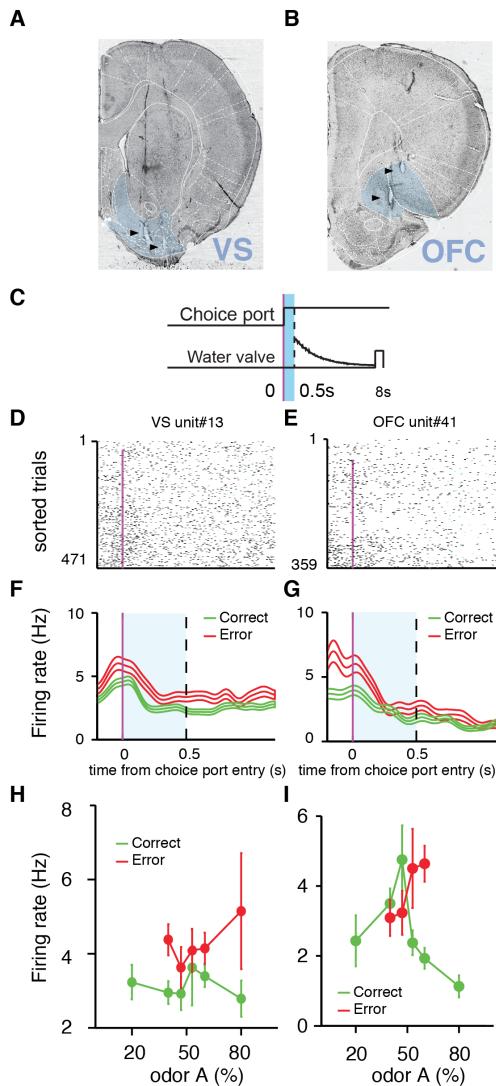


Figure 2.2 – Negative outcome predictive neurons in ventral striatum and orbitofrontal cortex

(A-B) - Example Nissl-stained coronal sections showing electrolytic lesion sites (black arrows) from tetrodes located in (A) ventral striatum (VS) and (B) orbitofrontal cortex (OFC).

(C) - Timing of outcome anticipation period. Neuronal activity was aligned to each trial choice port entry, signaled by the first break of photo-beams within each choice port, and analyzed for 500ms after choice port in – light blue bar.

(D-E) Activity of example neuronal units, from (D) VS and (E) OFC. Raster plots represent neural activity, with each row corresponding to a single trial and each tick mark to a spike. All the trials from a session are depicted sorted for the duration of movement from odor port to choice port. 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 (F) VS and (G) OFC example neuronal units (same as in D-E). Trials are grouped by trial outcome - correct (green) and error (red) PETHs. In PETHs, central lines are the firing rate averages and upper and lower lines represents s.e.m across trials.

(H-I) - Average firing rate of example (H) VS and (G) OFC example units (same as in D-E), during the outcome anticipation period, as a function of odor stimulus and trial outcome (correct, green; error, red). Error bars are s.e.m. across trials.

We next used receiver operator characteristic (ROC) analysis to define a measure of how well the firing rate of a neuron can be used to classify the outcome of a trial (Feierstein et al., 2006; Felsen and Mainen, 2008; Kepecs et al., 2008), See Experimental Procedures). For each neuron a outcome relative preference index was calculated, with negative values being assigned to cells with higher average firing rates for error trials (as in **Figure 2.2D-I**) and positive values for cells with higher firing rates for correct trials.

In the VS 25% of the recorded neurons (16/64) showed a significant outcome relative preference ($p<0.05$, permutation test) (**Figure 2.3A**). 75% of these outcome predictive cells (12/16) showed a negative outcome preference and the population average activity, as a function of outcome and odor mixture identity, revealed the “x-pattern” of *uncertainty* cells (**Figure 2.3C**). The remaining 25% (4/16) had positive outcome preference and *confidence* cells-like activity (**Figure 2.3D**). Choice accuracy varied with the firing rate of these populations - in the *uncertainty* neuronal population lower firing rates were associated with higher performance and higher firing rates to chance-level performance. The negative correlation between averaged activity and accuracy was significant (**Figure 2.3G**, $R=-0.83$ $p=0.02226$). On the other hand, a positive correlation between activity and accuracy was observed in the *confidence* population (**Figure 2.3H**, $R=0.95$ $p=0.00089$).

In the OFC 22% of the recorded neurons (13/59) showed significant outcome relative preference (**Figure 2.3B**), 62% of these (8/13) were *uncertainty* neurons (**Figure 2.3E**) and 38% (5/13) *confidence* neurons (**Figure 2.3F**). The activity of OFC *uncertainty* population was well correlated with accuracy (**Figure 2.3I**, $R=-0.89$ $p=0.00658$), but the correlation was weaker in the OFC *confidence* population (**Figure 2.3J**, $R=0.62$ $p=0.13663$).

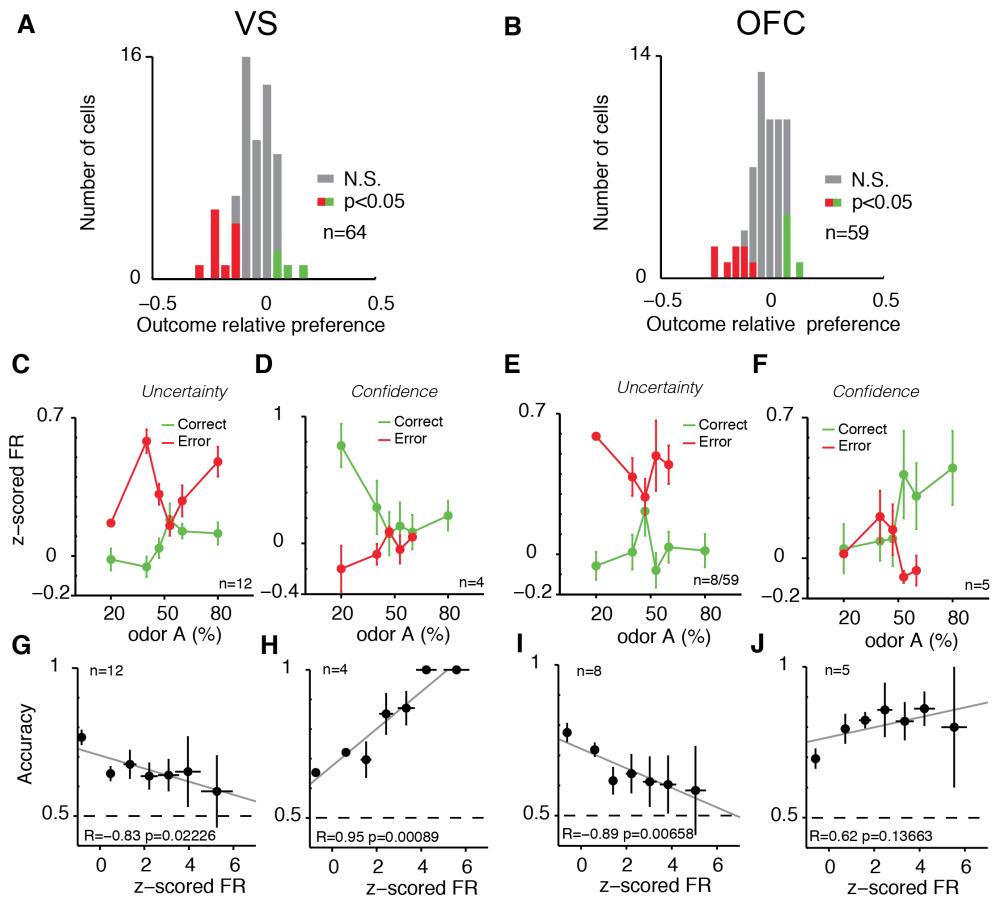


Figure 2.3 – Outcome predictive populations in ventral striatum and orbitofrontal cortex

(A-B) – Outcome relative preference for (A) VS and (B) OFC population of neurons during the outcome anticipation period. Outcome preference is calculated using ROC analysis (see Experimental Procedures). Color bars represent significant selectivity ($p < 0.05$, permutation test), green - positive outcome preference index (selective for correct trials), red - negative outcome preference index (selective for error trials) and gray – not significant.

(C-D) – Uncertainty and confidence cells in VS. Normalized firing rate of VS (C) uncertainty population (averages across negative outcome selective cells) and (D) confidence population (averages across positive outcome selective cells), as a function of odor stimulus and trial outcome (correct, green; error, red).

(E-F) – Uncertainty and confidence cells in OFC. Normalized firing rate of OFC (E) uncertainty population and (F) confidence population, as in C-D.

(G-H) – Mean choice accuracy as a function of the firing rate, for VS (G) uncertainty population and (H) confidence population. Normalized firing rates were binned for individual neurons and the mean accuracy was calculated for each range of firing rates, then averaged across cells. Gray line represents least-squares line. Correlation coefficients (R) and p values (p) for correlation are depicted.

(I-J) – Mean choice accuracy as a function of the firing rate, for OFC (I) uncertainty population and (J) confidence population, as in G-H. Error bars are s.e.m. across cells in C-J

Uncertainty and confidence populations predict waiting time

We next sought to investigate the role of *confidence* and *uncertainty* neuronal populations in the reporting of confidence via our waiting time post-wager. We have found that the activity of 31% (5/16) VS outcome predictive cells and 18% (2/11) OFC outcome predictive cells was also linearly correlated with waiting time during the outcome anticipation period (**Figure 2.S2**).

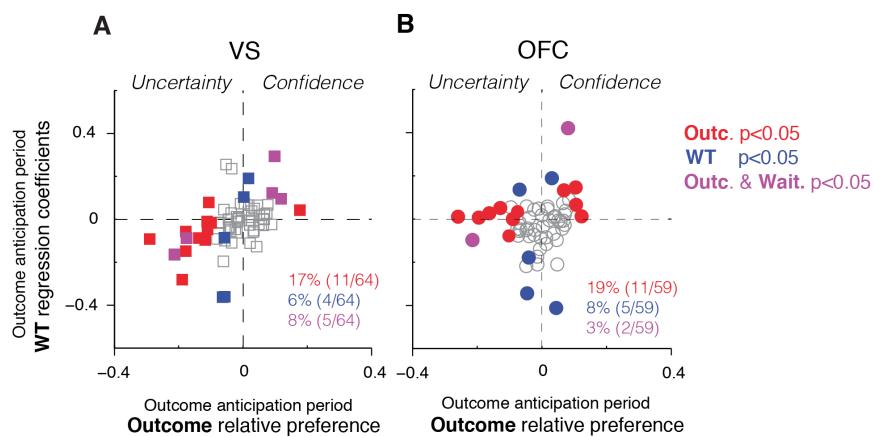


Figure 2.S2 – Uncertainty and confidence neurons in VS and OFC predict waiting time
(A-B) Population scatter plots of outcome relative preference and coefficient estimates for the linear regression of firing rate predicted by waiting time, during the reward anticipation period, for (A) VS and (B) OFC neurons. Each point is a single neuron, color-coded for significance ($p<0.05$) in the permutation test for the outcome relative preference (red) or in the linear regression (blue), or for both measures (magenta). In both VS and OFC significant cells for positive outcome relative preference are defined as confidence neurons and significant cells for negative outcome relative preference are defined as uncertainty neurons.

We've pooled VS and OFC *uncertainty* cells or VS and OFC *confidence* cells together to access the relationship between firing rates in *uncertainty* or *confidence* neuronal populations and our confidence report. In the uncertainty population (20/123 cells), higher average activity corresponded to shorter waiting times, and this negative correlation was significant (**Figure 2.4A**, $R=-0.79$ $p=0.03632$). The opposite was found to happen in the *confidence* population (9/123 cells) - on average when this population of neurons was more active rats

significantly waited longer for a reward (**Figure 2.4B**, $R=0.91$ $p=0.00412$).

To better understand the activity dynamics of these two populations, in relation with waiting time, we've generated population PETHs from trials ranked in intervals according to the distribution of waiting times (**Figure 2.4C-D**). Both *confidence* and *uncertainty* populations had a transient activation, locked to the time when the rat poked into the choice port, which was not clearly separable according to trial waiting time. Following this, activity was correlated with the waiting time percentile ranks either positively (*confidence*) or negatively (*uncertainty*), and this relation was only observed during this 500ms period. After this, and for at least 500ms, activity from the *uncertainty* population seemed to diverge into two binary categories – decreased activity for longer waiting times ($>60^{\text{th}}$ percentile) and sustained activity for shorter waiting times ($<60^{\text{th}}$ percentile) (Fig 2.4C) while activity of the *confidence* population converged and did not reflect waiting time (**Figure 2.4D**).

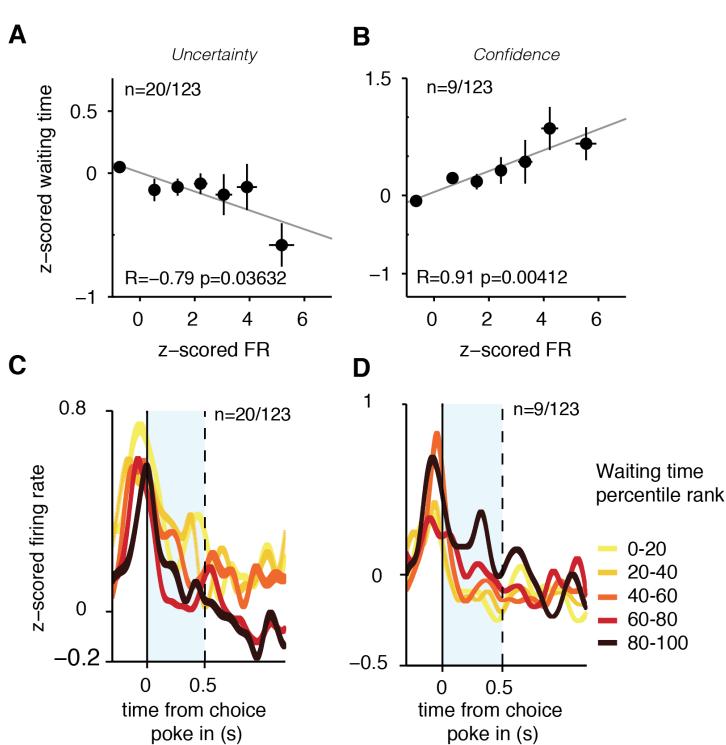


Figure 2.4 - Uncertainty and confidence neuronal populations predict waiting time

(A-B) - Mean normalized waiting time as a function of the firing rate, for total (A) uncertainty and (B) confidence recorded neurons. Normalized firing rates were binned for individual neurons, normalized averaged waiting time was calculated for each range of firing rates, and averaged across cells. Gray line represents least-squares line. Correlation coefficients (R) and p values (p) for correlation are depicted. Error bars are s.e.m. across cells.

(C-D) - Average PETHs of (C) uncertainty and (D) confidence recorded neurons. For each unit trials were grouped by waiting time percentile rank, from initial 20% (0-20, light yellow) to final 20% (80-100, dark brown). In light blue, the reward anticipation period. Lines are average normalized firing rates \pm s.e.m. across neurons.

Waiting time predictive neurons are different before, in the beginning or end of waiting

The population of neurons selected for significant outcome relative preference (*uncertainty* and *confidence*) during the outcome period had its activity correlated with waiting time at least at the beginning of waiting at the choice port. But it was unclear if those same cells could already be predicting waiting time (and outcome) before the rat had committed to a choice and poked into a choice port. Or, on the other hand, if the activity of these same cells kept being correlated with waiting time until the end of waiting and was when the leaving decision occurs. Given this, and to better understand the behavior of the waiting time predictive neurons as a population, we have looked into our 123 recorded neurons in two other trial periods: before waiting (activity aligned to choice port entry and analyzed for 500ms window before choice port in) and before leaving (activity aligned to choice port exit, and analyzed for a 500ms window before choice port out) and compared these two periods to the outcome anticipation period.

When comparing the period before choice port entry with the period after choice port entry 9% of the cells (10/123) could predict waiting time only after choice port in, 5% (6/123) could predict waiting time only before choice port in, and 5% (6/123) could predict waiting time during both periods (**Figure 2.5A**).

The amount of outcome predictive cells between these two periods was similar- 14% (17/123) were predictive of outcome in only one of the periods and 10% (12/123) could predict outcome during both periods (**Figure 2.S3A**).

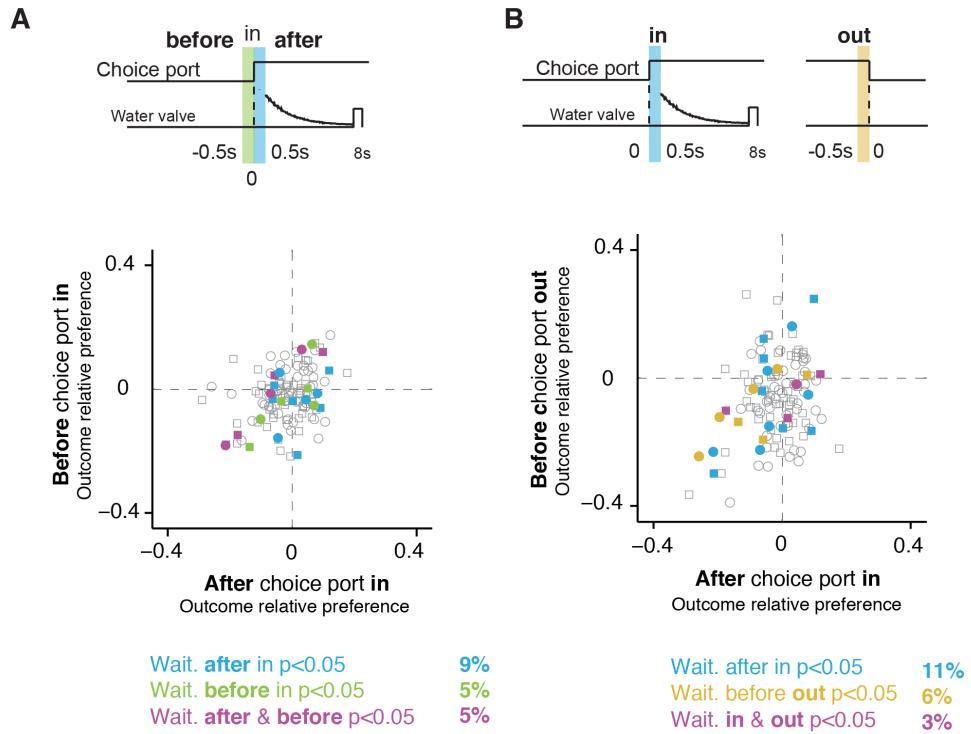


Figure 2.5 – Waiting time predictive neurons are different before waiting, in its beginning or end.

(A) – (top) Neuronal activity was aligned to each trial choice port entry and analyzed for 500ms after choice port in –blue bar, or aligned to each trial choice port entry and analyzed for 500ms before choice port in – green bar

(bottom) Population scatter plot of outcome relative preference during a 500ms window after choice port in (reward anticipation period), and outcome relative preference during a 500ms window before choice port in (movement period). Each point corresponds to a cell, color-coded for significance ($p<0.05$) in a linear regression of firing rate with waiting time as predictor (blue, waiting time regression after choice port in; green, before choice in and magenta, during both periods). Unfilled points are non-significant cells. Squares correspond to VS cells, circles to OFC cells.

(B)– (top) Neuronal activity was aligned to each trial choice port entry and analyzed for 500ms after choice port in –blue bar, or aligned to each trial choice port exit and analyzed for 500ms before choice port out – yellow bar.

(bottom) Population scatter plot of outcome relative preference during a 500ms window after choice port in (reward anticipation period), and outcome relative preference during a 500ms window before choice port out. Each point corresponds to a cell, color-coded for significance ($p<0.05$) in a linear regression of firing rate with waiting time as predictor (blue, waiting time regression after choice port in; yellow, before choice port in and magenta, during both periods). Unfilled points are non-significant cells. Squares correspond to VS cells, circles to OFC cells.

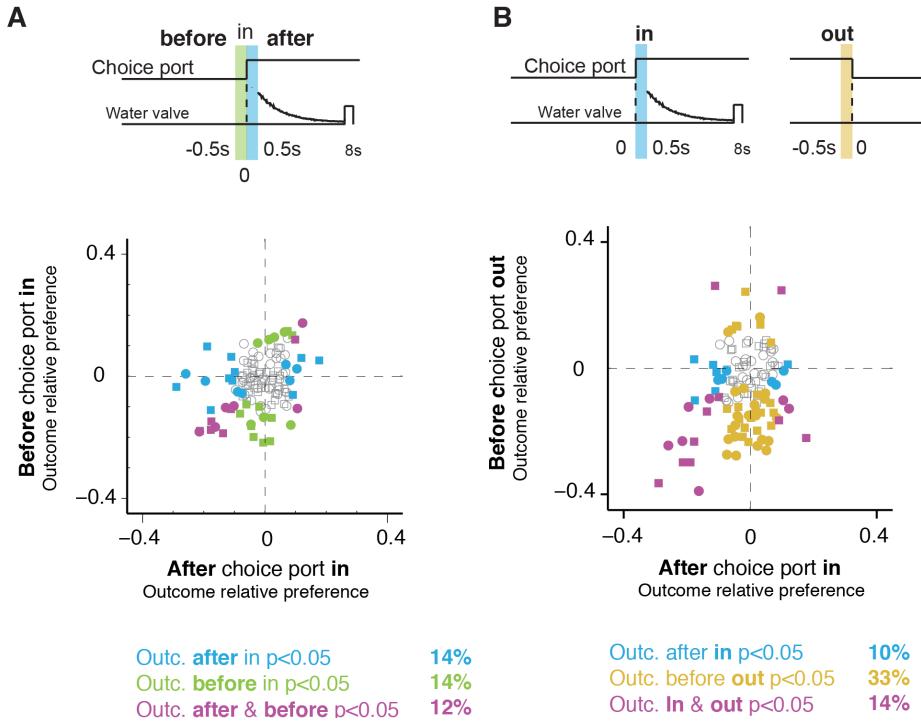


Figure 2.S3 – Outcome predictive before waiting, in its beginning or end.

(A) - (top) Neuronal activity was aligned to each trial choice port entry and analyzed for 500ms after choice port in –blue bar (left panel), or aligned to each trial choice port entry and analyzed for 500ms before choice port in – green bar.

(bottom) Population scatter plot of outcome relative preference during a 500ms window after choice port in (reward anticipation period), and outcome relative preference during a 500ms window before choice port in (movement period). Each point corresponds to a cell, color-coded for significance ($p<0.05$) in the permutation test for the outcome relative preference (blue, after choice port in; green, before choice in out and magenta, during both periods). Unfilled points are non-significant cells. Squares correspond to VS cells, circles to OFC cells.

(B) - (top) Neuronal activity was aligned to each trial choice port entry and analyzed for 500ms after choice port in –blue bar (left panel), or aligned to each trial choice port exit and analyzed for 500ms before choice port out – yellow bar.

(bottom) Population scatter plot of outcome relative preference during a 500ms window after choice port in (reward anticipation period), and outcome relative preference during a 500ms window before choice port out. Each point corresponds to a cell, color-coded for significance ($p<0.05$) in the permutation test for the outcome relative preference (blue, after choice port in; green, before choice port out and magenta, during both periods). Unfilled points are non-significant cells. Squares correspond to VS cells, circles to OFC cells.

Moreover, the outcome relative preference indices of waiting time predictive cells were dispersed all over the outcome relative preference distributions, in both periods, with the occurrence of cells which could predict waiting time but not outcome (**Figure 2.S4A**).

When comparing the period after choice port in with the period just before choice port out 11% of the cells (13/123) predicted waiting time only during before choice port in, 6% (7/123) only before choice port out and 3% (4/123) would be common for both periods (**Figure 2.5B**).

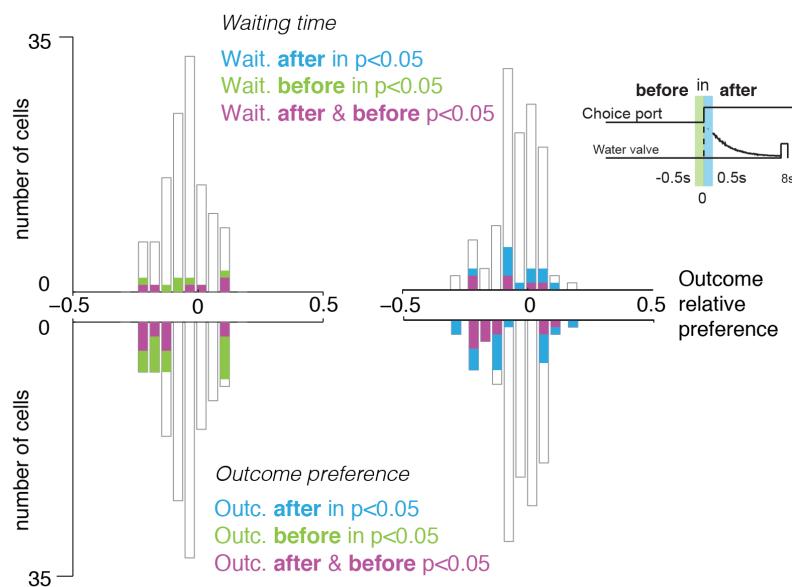
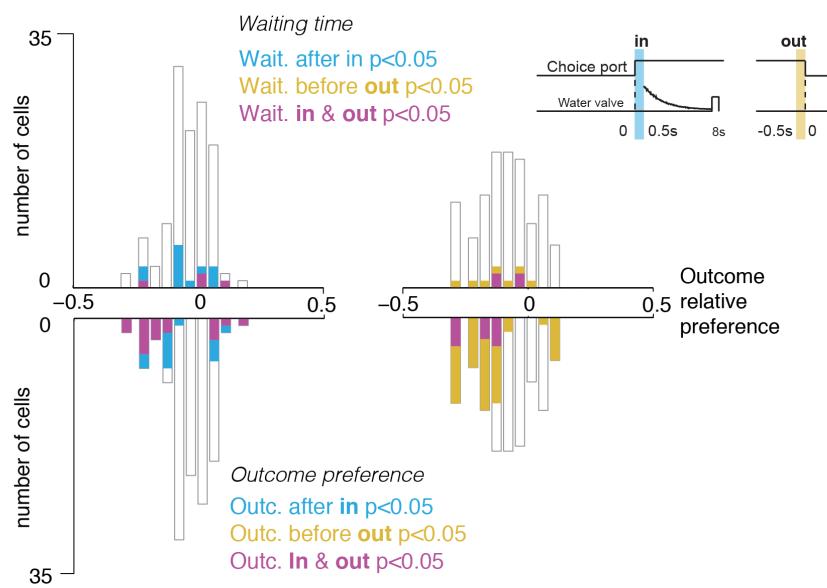
A**B**

Figure 2.S4 – Population of neurons predictive of outcome or waiting time, before waiting, in it's beginning or end.

(A)- Histogram of outcome relative preference indices for neuronal activity aligned to trial choice port entry and analyzed for 500ms before choice port in (left panel), or analyzed for 500ms after choice port (right panel). Histograms are color coded for significance for two parameters, based on a linear regression of firing rate with waiting time as predictor (top histograms), or based on the outcome relative preference permutation test (bottom histograms). Green indicates number of cells significant during the 500ms before choice port in, blue, during the 500ms after choice port in and magenta during both periods.

(B) - Histogram of outcome relative preference indices for neuronal activity aligned to trial choice port entry and analyzed for 500ms after choice port in (left panel), or for neuronal activity aligned to trial choice port exit, analyzed for 500ms before choice out (right panel). Histograms are color coded for significance for two parameters, based on a linear regression of firing rate with waiting time as predictor (top histograms), or based on the outcome relative preference permutation test (bottom histograms). Yellow indicates number of cells significant during the 500ms before choice port out, blue, during the 500ms after choice port in and magenta during both periods.

There was a difference in the amount of outcome related cells between these two periods. Just before choice port out rats should have already experienced the presence (or absence) of outcome. Maybe due to this 33% of the cells (40/123) had a significant outcome relative index only during this period (**Figure 2.S3B**). From these 83% (33/40) had a negative outcome preference index and thus signaled errors. Just after choice port in 14% of the cells (17/123) could predict outcome and also be sensitive to it's absence or presence just after choice port out, and 29% of these cells (5/17) inverted their outcome relative preference signal from beginning to end of waiting. The outcome relative preference indices of waiting time predictive cells were also dispersed all over the outcome relative preference distributions, in both periods, with the occurrence of waiting time predictive cells which could not predict or distinguish between outcomes (**Figure 2.S4B**).

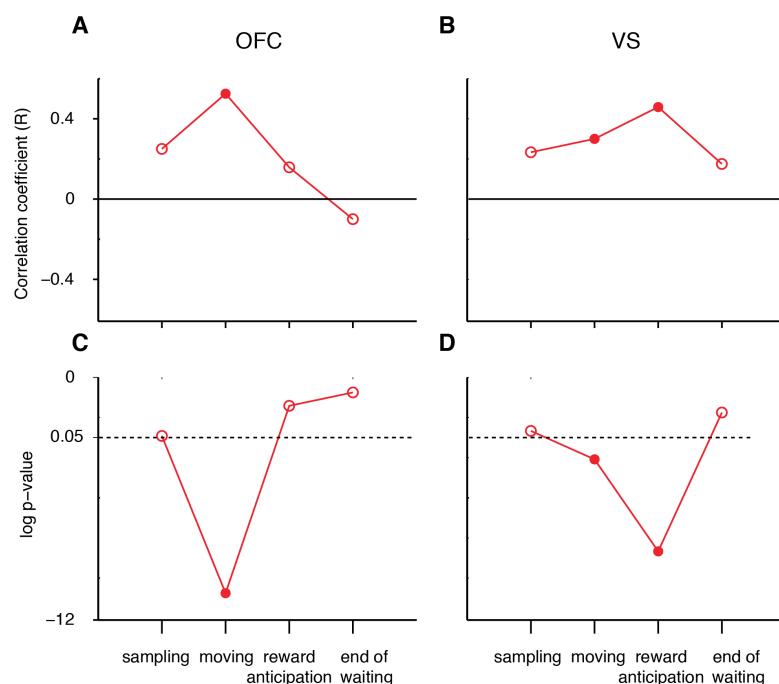


Figure 2.S5 – Within trial dynamics of waiting time and decision confidence correlations

(A,B) – Correlation coefficients of outcome relative preference and waiting time regression coefficients for (A) VS neurons and (B) OFC neurons, during four periods of a trial: sampling - activity aligned at odor port out and analyzed for 200ms before odor port exit; moving - activity aligned at choice port in and analyzed for 500ms before choice port entry; reward anticipation - activity aligned at choice port in and analyzed for 500ms after choice port entry; and end of waiting - activity aligned at choice port exit and analyzed for 500ms before choice port out. Filled circles, $p<0.05$

(C,D) – p values for correlation of outcome relative preference and waiting time regression coefficients for (C) VS neurons and (D) OFC neurons, during the above mentioned four periods of a trial. Filled circles, $p < 0.05$.

Overall, the population of OFC neurons had their outcome relative preference indices significantly correlated with their waiting time linear regression coefficients (**Figure 2.S5A, C**) in the 500ms window before choice port entry ($R=0.52$, $p=2.2 \times 10^{-5}$), but not during reward outcome anticipation ($R=0.16$, $p=0.234$) or in the end of waiting ($R=-0.1$, $p=0.454$). The population of VS neurons had their outcome relative preference indices significantly correlated with their waiting time linear regression coefficients (**Figure 2.S5B, D**) in the 500ms window before choice port entry ($R=0.45$, $p=0.0002$), and during reward outcome anticipation ($R=0.3$, $p=0.0.016$) but not in the end of waiting ($R=0.17$, $p=0.1752$). We have also analyzed a shorter 250ms window of activity before odor port exit, correspondent to the time animals are sampling the odor mixture. Before decision formation, both OFC and VS population of neurons did not have their outcome relative preference indices significantly correlated with their waiting time linear regression coefficients (OFC, $R=0.25$, $p=0.0544$; VS, $R=0.23$, $p=0.0692$).

DISCUSSION

To explore the neural correlates of confidence, uncertainty and its behavioral report (waiting time) we recorded electrophysiological activity of single neurons from the orbitofrontal cortex (OFC) and ventral striatum (VS) of behaving animals, performing a odor-mixture categorization task with a time wager report of confidence. Just after a choice was made we have found, in both the OFC and VS, populations of neurons that computed *confidence* and *uncertainty* (**Figure 3.3**) and that could well predict the time animals were willing to wait for a reward (**Figure 3.4**).

While the OFC was previously found to have neuronal correlates of uncertainty (and confidence) (Kepecs et al., 2008; O'Neill and Schultz, 2013) the relation between this activity and a behavioral report of

decision confidence was yet to be studied. Regarding VS, to our knowledge, this is the first report of single neuron correlates of decision confidence in this area although prior human functional imaging studies (Daniel and Pollmann, 2012b; Grinband et al., 2006; Hebart et al., 2014a; Preuschoff et al., 2006) already hypothesized a role for the VS in decision making under uncertainty. Most notably, in Hebart et al, VS was the primary candidate region identified for encoding perceptual confidence, amongst frontal cortical regions.

In our hands the total fraction of outcome predictive cells during the reward anticipation period was similar in both OFC and VS, and their activity was well correlated with decision accuracy. As a first approach analysis we decided to combine *uncertainty* and *confidence* populations from both regions and have observed that higher firing rates in *uncertainty* cells predicted shorter waiting times and the opposite was true for *confidence* cells. This correlation between a trial-by-trial graduate confidence report and activity of cells computing perceptual decision confidence/uncertainty was only previously shown in the primate supplementary eye field (Middlebrooks and Sommer, 2012).

When looking closer to firing rate dynamics of these two populations we saw different tiled activity depending on the time animals waited for reward, but this difference was mainly observed in the reward anticipation period - most likely reflecting our selection criteria of outcome predictive cells. Those dynamics did not resemble a ramp-to-threshold pattern, expected if these populations were responsible for accumulating confidence-based evidence that would result in the leaving decision upon reaching a threshold. It is possible that these populations of cells are instead feeding a accumulator network, located in one other brain region, which would take into account the amount of decision confidence, amongst other variables, and ultimately lead to spontaneous leaving decisions. The rodent pre-motor cortex would be a candidate area for such a role (Murakami et al., 2014). Also, the fact that in the *confidence* population the tiled difference collapsed soon after 500ms from choice port entry might indicate that these cells are not keeping

track of time spent waiting. The activity of the *uncertainty* population seemed to collapse into a binary pattern, no longer spread across the waiting time distribution but separated for trials <60% or >60% of waiting time distribution. This either reflects a difference between expected outcomes that does not correlate with confidence report from that time point on, or a change to a binary confidence encoding (Martinez-Garcia et al, 2014). Both populations had a peak activity just before choice entry, which also did not clearly reflect waiting time.

The fact that waiting time tiled activity only held during the outcome preference selection window lead us to explore the idea that the population of outcome predictive cells, and confidence-report correlated cells, was not homogenous across the trial. We had chosen to analyze activity specifically during the outcome anticipation period, following a previous report (Kepecs et al. 2008), which had found neural correlates for confidence in OFC during that epoch. This period seemed ideal since the animals had already committed to a choice but were not yet rewarded. Nonetheless we looked at two other trial epochs, 1) during 500ms before choice port entry, when animal was moving towards the choice port, we expected that some neurons would already be carrying a decision confidence estimation maybe correlated with waiting time; and 2) for 500ms before leaving choice port, where animals should already be aware of the presence of reward in correct trials, or it's absence in error or catch trials.

For this analysis it was important to go back to individual neurons, and correlate single unit activity with waiting time or outcome preference. The fraction and identity of outcome predictive or waiting time predictive single neurons changed across different trial epochs, revealing interesting within-trial dynamics of the confidence signal. The highest fraction of cells with activity correlated with waiting time was observed in the reward anticipation period, but there was still a considerable amount of cells which were predicting waiting time exclusively just before entering the choice port or just before leaving it. This within-trial span of waiting time correlated activity might strengthen the idea that

these neurons feed an integrator of waiting time, in different stages of the accumulation process, as suggested before.

Just after the odor guided-decision had taken place, and while animals were moving towards a choice port, there was already a small fraction of cells with significant outcome relative preference, either limited to that period or with sustained significant preference spanning until the reward anticipation period. This unravels that decision confidence estimates start to be computed early on in OFC and VS neurons. The amount of neurons distinguishing error trial from correct trials between just after choice port in and just before choice port out is quite different. A big proportion of the recorded neurons significantly increase their activity after an error was made, in what could be revealing computations necessary for detecting absence of reward. Although the aim of this study is other, it is widely reported that OFC and VS are involved in these processes (eg. Schultz et al., 2000). More so, a good proportion of these neurons shifted their preference from preferring correct choices, during the reward anticipation period, to preferring error choices, in what might be a signal for the discrepancy between expected outcome and actual outcome – prediction error – something previously observed in VS and OFC functional imaging (Doherty et al., 2003).

Overall, the confidence report computation is dynamically shaped by the decision confidence/uncertainty signal, within the OFC-VS circuitry, throughout the trial starting just after decision. The outcome relative preferences of neurons in VS and OFC significantly correlate with the trial waiting times during the movement period. During this period OFC outcome predictive activity is more strongly correlated with the confidence report. After this, and while animals are anticipating the reward the correlation is significant only in the VS neurons but the significance is lost just before animals leave the choice port to start a new trial. Previously (Lak et al., 2014) OFC pharmacological inactivation was found to disrupt the waiting time dependency on confidence, without affecting perceptual decision making, or mean total waiting time. This disruption, although significant, was not total. In line with our results it

might be feasible to think that 1) OFC inactivation affected mostly the decision confidence report by having diminished outcome relative preference related activity just after decision formation and 2) even after OFC inactivation, VS outcome relative preference related activity might be responsible for keeping a waiting time behavior still partially dependent on decision confidence.

We have shown that the neurons OFC-VS circuitry can play a role in guiding post-wager decisions based in subjective beliefs of decision confidence. These two regions are thought to be important in model-based decision making (McDannald et al., 2011) or serve the role of the *critic* in actor-critic implementation of reinforcement learning (Botvinick et al., 2009), to track progression through a task sequence (Shidara et al., 1998) and act in conjunction to influence actions towards outcomes (Simmons et al., 2007). Here we propose that the orbitofrontal-ventral striatal circuitry is also central for using metacognitive signals, such as estimates of decision confidence, to shape adaptative behaviors in an uncertain world.

INTERPLAY BETWEEN CONFIDENCE AND VALUE TO GENERATE A CONFIDENCE REPORT

SUMMARY

Perceptual categorical decisions take place based on stimulus information and reinforcement-related factors, such as the value of outcomes. Before outcome is known, while animals are waiting for a reward, the brain is able to compute an evaluation of how likely it is that the decision just made was a correct one. If given the chance, they can adapt their behavior accordingly, and for instance, choose to give up waiting and start again. This behavior is, like a metacognitive judgment, based on decision confidence estimates.

When a decision is uncertain animals bias their choices in favor of the most rewarding option. But it is not clear if reward magnitude affects metacognitive judgments. To better understand the effect of reward in post-decision confidence judgments we have trained rats to perform a modified version of the waiting task, which had a block-wise reward manipulation. We have observed that the reward manipulation biased animals' choices, affected their performance but did not seem to alter the waiting time confidence report. By devising a SDT model which assumed a dual-route processing of confidence we were able to explain our behavioural results.

INTRODUCTION

Animal lives are punctuated by decisions. Uncertainty and reward value play a major role in guiding and influencing each decision made - errors are more prevalent upon high uncertainty conditions and choices are more abundant towards most valuable options. Imagine a stockbroker analyzing highly complex and uncertain financial information. In doubt he opts to invest towards a very valuable option. Before he actually knows the outcome of his decision he can estimate his degree of confidence and sell back his shares when confidence is lacking. But how is his confidence affected by the magnitude of expected reward? Is his judgment affected by greed?

Human and non-human animals have the ability to report “metacognitive” confidence judgements on the quality of a decision (eg. Kepecs and Mainen, 2012; Smith et al., 2003; Yeung and Summerfield, 2012). The importance of understanding the extent to which these confidence judgements reliably predict decision accuracy -calibration- has long been studied in human decision making (Baranski and Petrusic, 1994; Griffin, 1993; Harvey, 1997). Through this effort it was observed that decision confidence was often uncalibrated, with a propensity for overconfident estimations following hard decisions. Later on individual differences in confidence calibration were linked to anatomical differences in prefrontal cortex areas (Fleming et al., 2010). Furthermore, it was proposed (Johnson and Fowler, 2011) that holding incorrect beliefs about one’s own capability could actually be a beneficial strategy in everyday life, depending on small changes in benefit/cost ratio of decisions, with advantage for overconfident individuals when deciding under high levels of uncertainty.

Despite all previous work it is not well understood how is a confidence report affected by unbalanced reward benefits. Recently, in a attempt to better understand how subjective confidence and valuation of choice options interact, De Martino, et al, (2013) observed a separation between the two variables, confirming the notion that high confidence could be reported after a low-value choice and vice versa. In this study human subjects were asked to make value-based decisions and accuracy was as measure of whether decisions were following an *a posteriori* defined subjective preference (Becker-DeGroot-Marschak mechanism). This task design meant that correct or wrong decisions were defined by individual subjective preferences. It was left unknown how is the interplay between value and confidence when the external world defines the rule for decisions.

In reward based decisions the amount of offered reward biases choices towards the most valuable options (Lau and Glimcher, 2008; Roesch et al., 2006; Sugrue et al., 2004). The same can be observed in perceptual decisions, with the amount of choice bias depending on the sensory cue

contrast (Navalpakkam et al., 2010; Wang et al., 2013). It has been previously reported that animals can compute decision confidence estimates of perceptual decisions, and derive future behaviors accordingly (Kepcs et al, 2008; Kiani & Shadlen, 2009; Komura, et al, 2013; Lak et al, 2014). The confidence-driven behaviors were highly correlated with the difficulty of the task (degree of uncertainty) and expected outcome (probability of being correct), but the effect of different offered reward amounts in a confidence report for perceptual decisions remained to be determined.

Signal Detection Theory (SDT) models can be used to generate predictions that account for choice behavior and confidence estimation, in perceptually driven decisions under uncertainty (Deco et al., 2013; Hebart et al., 2014b; Kepecs and Mainen, 2012; Kepecs et al., 2008). These models have the caveat that do not reflect a temporal dimension but can be particularly useful and simple to apply in tasks where the time for sensory evidence accumulation does not seem to be a function of evidence contrast, such as in odor-guided categorizations (Uchida and Mainen, 2003; Zariwala et al., 2013). In tasks where the dynamic integration of evidence plays a important role in shaping response times, integrator or accumulator models can be used to derive choice and confidence ((Kiani and Shadlen, 2009; Vickers and Packer, 1982) Both types of models postulate that confidence estimation is a intrinsic property of the decision making process. In order to compute behaviors which monitor performance, such has confidence judgements, it was proposed that a second layer of processing which has access to confidence estimates (Charles et al., 2014; Cul et al., 2009; Insabato et al., 2010a; De Martino et al., 2013) might be implemented by the brain.

To better understand the effect of reward in post-decision confidence judgments we have trained rats to perform a two-odor guided categorization task, which includes a post-wager decision confidence report and a block-wise reward manipulation. We have observed that the reward manipulation biased animals' choices, affected their performance but did not seem to alter the waiting time confidence report. By devising

a SDT model that assumed a dual-route processing of confidence we were able to predict, and explain, our behavioural results.

RESULTS

Odor categorization task with confidence report and reward manipulation

We have trained rats ($n=6$) to perform a modified version of the 2AFC odor categorization task (Uchida and Mainen, 2003), which incorporates a post-decision wager (Kepecs and Mainen, 2012; Lak et al., 2014). In this paradigm each trial consists of two different decisions, a perceptual decision towards one of two choice ports and a post-decision wager - waiting at the choice port - for incorrect trials and a fraction of correct trials where reward is omitted, catch trials (Figure 3.1A, See Experimental Procedures II & III). This post-decision wager is based on the willingness to wait for a reward, where the amount of time waited is a function of decision confidence and the opportunity cost of waiting. Waiting time (WT) will be higher for higher levels of confidence and lower costs of waiting (Lak et al., 2014).

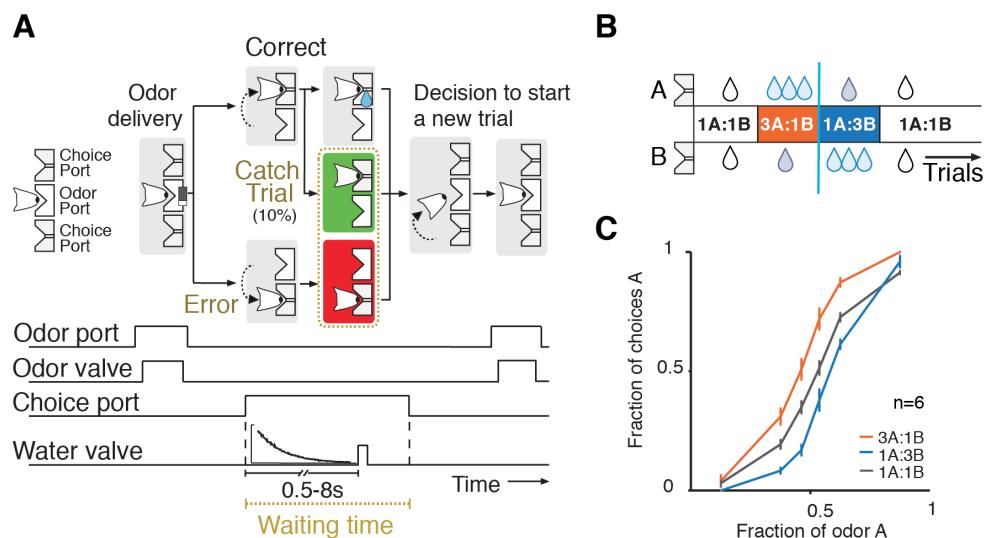


Figure 3.1 – Odor categorization task with confidence report and reward manipulation

(A) - Schematic of the behavioral paradigm. To start a trial, rats entered the central odor port and after a pseudorandom delay of 0.2–0.5 s a mixture of odors was delivered. Rats moved to

one of the lateral choice ports and in correct trials they waited for a drop of water to be delivered after a pseudorandom delay, drawn from a exponential distribution with decay of 1.5s, 0.5s offset and 8s maximum. In error trials no water was delivered, and so was the case for a small percentage of correct trials – catch trials. The time rats were willing to wait with their snout in the reward port, in both error trials and catch trials – waiting time – was considered to be the confidence report for this task. In these trials rats were able to start a new trial 1s after leaving the reward port. **(B)** - Reward magnitude manipulation. Each session starts with a block (1A:1B) of 120-140 trials where both choice ports deliver the same amount of water (0.024ml). This is followed by two reward manipulation blocks – a 3A:1B block of 120-140 trials where one of the choice ports A delivers three times more water (0.072ml) than the opposite port B and then this relation is reversed for the next set of 120-140 trials (1A:3B). For the remaining trials of the same session the choice ports deliver again the same amount of reward (0.024ml) (1A:1B). The initial most rewarded side alternates from session to session. **(C)** - Effect of reward magnitude manipulation on performance. Psychometric functions of choices in 1A:1B (gray), 3A:1B (orange) and 1A:3B (blue) blocks. Values depicted are averages for a population of 6 animals. Error bars are s.e.m.

Classically the difficulty of this task is manipulated by varying the contrast of the odor mixture, pseudo-randomly delivered in a central port (6%, 20% and 60% contrasts, corresponding to the odor ratios 20:80, 40:60, 47:53 and their counterparts, 80:20, etc.) and so accuracy increases with mixture contrast (Uchida and Mainen, 2003). To study the interplay between reward value and confidence we introduced a reward magnitude manipulation in a block-wise design (**Figure 3.1B**). During a session rats experienced an initial block of trials where water was evenly delivered at the choice ports (0.024ml, 1A:1B block). This was followed by another block of trials where in one choice port the amount of reward delivered was the same as previously (0.024ml) and in the opposite choice port rats could obtain three times more water (0.074ml). In the next block of trials the unbalance in the magnitude of water delivery was reversed amongst ports. Depending of which port, A or B, was more rewarding we denominated these blocks as 3A:1B or 1B:3A. After these two blocks and until the end of the session the water delivery was again balanced (1A:1B).

As previously observed in a similar reward manipulation task (Wang et al., 2013), rats choices were biased towards the most rewarded side in a difficulty dependent manner (**Figure 3.1C**) - biases are observed, mostly for when the odor mixture was closer to the categorization boundary (50:50).

Reward manipulation biases choices and affects accuracy

We further analyzed the effect of our block-wise reward manipulation design on choices and accuracy. Regardless of stimulus contrast we looked at sliding average choices relative to which of the choice ports the animals decided to move to - towards side A or side B (**Figure 3.2A top, middle**). To obtain a measure of choices according to their expected reward magnitude we divided the trials into 1:1 (equally rewarding), 1:3 (less rewarding) or 3:1 (more rewarding) choices and averaged across the population of animals (**Figure 3.2A bottom**). During the initial 1A:1B block choices are balanced, occurring around 50% of the times to each port. This would be expected due to our pseudorandom schedule of stimulus delivery - the probability of delivered stimulus cueing a choice to each port was set to be even (0.5 for both sides), for a blocks of 40 trials. Upon introducing an unbalance in the amount of delivered reward animals started biasing their choices towards the most rewarded choice port (3:1) away from the least rewarded choice port (1:3). Reversing the reward magnitude bias also reversed the bias in animals' choices. An opposite effect could be observed in the dynamics of accuracy (**Figure 3.2B top, middle**).

In sum, reward manipulation biased choices (**Figure 3.2A bottom**, $p<0.05$, Wilcoxon signed rank test, for 1:1 and 1:3 or 1:3 and 3:1 comparisons, across $n=6$ rats). On evenly rewarded trials (1:1) animals chose to go to side B in $55\% \pm 1.7\%$ of the choices. Due to the reward magnitude manipulation animals chosen $61\% \pm 1\%$ of the times the most rewarded choice port (3:1) and $39\% \pm 1\%$ the least rewarded choice port (1:3).

Reward manipulation affected performance (**Figure 3.2B bottom**, $p<0.05$, Wilcoxon signed rank test, for 1:1 and 1:3 or 1:3 and 3:1 comparisons, across $n=6$ rats). Animals' average accuracy was $76\% \pm 1\%$ on the choices made during the evenly rewarded blocks. In choices were animals followed the reward magnitude bias, and went towards the most rewarded side, accuracy slightly dropped to $73\% \pm 1\%$. On the opposite

direction, when we look at 1:3 choices, these were on average $84\% \pm 1\%$ accurate, higher than in 1:1 and 3:1 choices.

Our reward manipulation biased choices towards the most rewarded side. Also rats performed better when counteracting the biasing force that an offer for a higher amount of reward provided. What was the effect of the reward manipulation on the confidence report – waiting time?

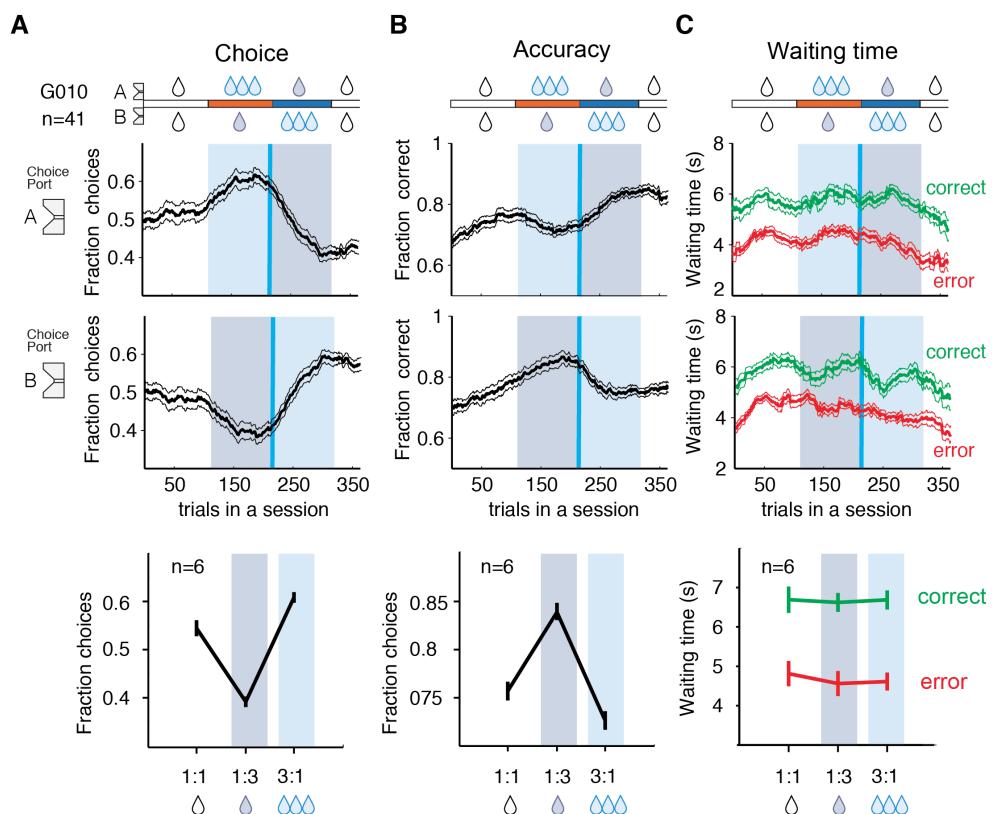


Figure 3.2 – Effect of reward magnitude manipulation on choices, accuracy and confidence report.

(A) - Effect of reward manipulation on choice. Average choice behavior of one animal (*top*) towards port A, or (*middle*) towards port B. Vertical blue line marks mean boundaries of 3A:1B to 1A:3B block transitions. Blue and purple boxes mark limits of reward manipulation blocks. Black thick and thin lines are 24 trial sliding averages and s.e.m, across 41 sessions. (*bottom*) Average fraction choices towards the least rewarded port (1:3 – purple box), towards the most rewarded port (3:1 – blue box), and towards port B in 1:1 trials. Values depicted are averages for a population of 6 animals. Error bars are s.e.m.

(B) - Effect of reward manipulation on accuracy. Accuracy on choices (*top*) towards port A or (*middle*) port B for one animal. (*bottom*) - accuracy for 1:1, 1:3 or 3:1 choices for n=6 animals.

(C) - Effect of reward manipulation on waiting time. Waiting time for correct catch (green) or error (red) trials on choices towards (*top*) port A or (*middle*) port B for one animal. (*bottom*) - waiting time for 1:1, 1:3 or 3:1 choices for n=6 animals.

Reward manipulation does not affect waiting time

If we hypothesize that animals wait for water depending on the amount of rewards they expect to obtain we would be led to predict that animals would be willing to spend more time waiting for a bigger reward (3:1) and less time waiting for a less rewarding choice (1:3). In this case waiting time would assume a pattern similar to the bias of reward manipulation on choices (**Figure 3.2A bottom**). An alternative hypothesis would be that waiting time is mostly related to the probability of being correct, hence mostly related with confidence. Given this, waiting time would be higher for when animals expect to be more accurate, and waiting time would assume a pattern similar to what is observed for the effect of reward manipulation on accuracy (Fig 3.2B *bottom*). Previous work has modeled waiting time as being positively correlated with confidence and negatively correlated with the cost of waiting (Lak et al., 2014). So, if the cost was to be (negatively) related to the expected reward value, then we might foresee an interaction between both previous hypotheses.

When we further analyzed the within-session dynamics of waiting time for individual animals we could not identify a pattern of dependency of waiting time on the different magnitudes of reward available in each port (**Figure 3.2C top and middle**). This is better revealed when we analyze the average waiting time for the population of animals. Rats wait on average a similar amount of time for higher and lower amount of reward, both for correct or error trials. This is no different from the time they are willing to wait in the 1:1 condition ($p>0.05$, Wilcoxon signed rank test, for all comparisons, across $n=6$ rats). The reward manipulation does not affect waiting time, and the amount of evidence available for each decision does not influence this result (two-way Anova, $p<0.001$ for the effect of odor contrast on waiting time, $p>0.05$ for effect of value and for the interaction, in either correct or error choices), (**Figure 3.S1A**).

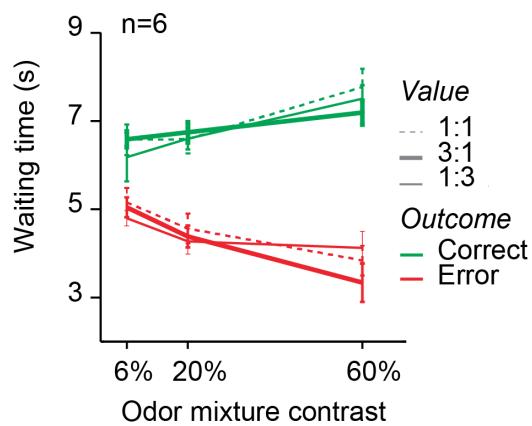
A

Figure 3.S1 – Confidence report odor contrast and reward magnitude

(A) Waiting time as a function of stimulus contrast, expected outcome and expected reward amount. Waiting times for the three different odor mixture contrasts, for correct (green) or error (red) choices, and 1:1 (dashed line), 3:1 (thick line) or 1:3 (thin line) choices.

Values are means and error bars s.e.m. across a population of 6 animals.

We also investigated the relation between waiting time and expected reward magnitude by analyzing single animal average waiting time, for different choice reward magnitudes and different expected outcomes (**Figure 3.S2A**). We could distinguish a clear separation between waiting times for correct and error choices, with animals waiting longer when expecting a trial to be correct, for (6/6 rats, $p < 10^{-11}$, Wilcoxon signed rank test). Animals didn't wait differently in 1:3 compared to 3:1 choices (0/6 rats during error trials, 1/6 rats during correct choices, $p < 0.05$, Wilcoxon signed rank test).

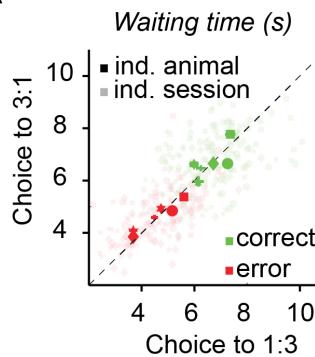
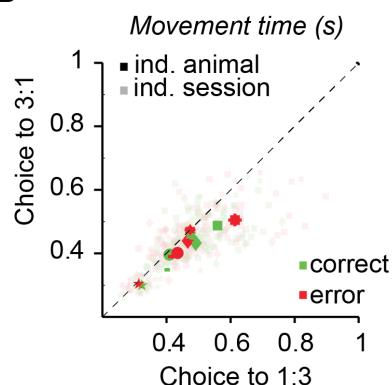
A

Figure 3.S1 – Waiting time, movement time and reward magnitude

(A) - Scatter plot of waiting time for 1:3 choices and 3:1 choices. Each animal average waiting time, for correct catch (green) and error (red) is plotted, with a unique marker type per animal. Each individual session average waiting time, for correct catch (light green) or error (light red) trials is also plotted. Error bars are s.e.m., and the markers hide most of them.

(B) - Scatter plot of movement time (defined as the interval between the time of leaving the odor port and the time of entry into a choice port), for 1:3 choices and 3:1 choices. Each animal average movement time, for correct (green) and error (red) is plotted, with a unique marker type per animal. Each individual session average movement time, for correct (light green) or error (light red) trials is also plotted. Error bars are s.e.m., and the markers hide most of them.

B

In contrast, movement time -the time that rats take to move from odor port to the choice port – seems to be a good predictor of the amount of reward rats are expecting to obtain (3/6 rats during error trials, 5/6 rats during correct choices, $p<0.05$, Wilcoxon signed rank test), and a fair predictor of outcome (3/6 rats, $p<0.05$, Wilcoxon signed rank test), (**Figure 3.S2B**).

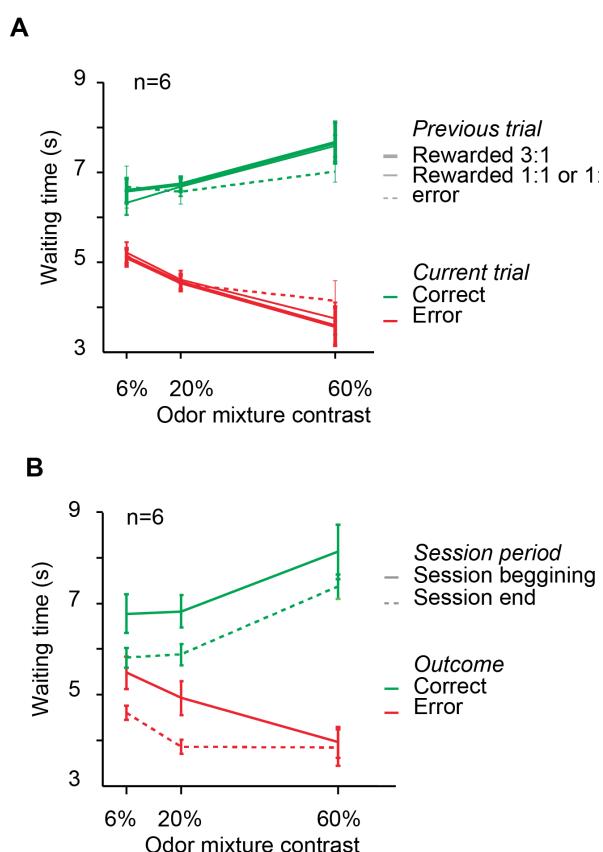


Figure 3.S3 – Confidence report and trial history

(A) Waiting time as a function of stimulus contrast and expected outcome, for different previous trial outcomes. Waiting times for the three different odor mixture contrasts, for correct (green) or error (red) choices, depending on whether the previous trial was a error trial (dashed line), a correct rewarded trial in 1:1 and 1:3 choices, or a correct rewarded trial for 3:1 choices.

(B) Waiting time as a function of stimulus contrast and expected outcome, for different session periods. Waiting times for the three different odor mixture contrasts, for correct (green) or error (red) choices, depending on whether the trials were done in the first half (solid line) or second half (dashed line) of the session. Averages and error bars (s.e.m.) across a population of 6 animals.

We have also observed that previous trial outcome history did not seem to influence waiting time – the fact that animals experienced 0.024ml of water or three times that amount in the previous trial did not change the following trial waiting time (two-way Anova, $p<0.001$ for the effect of odor contrast on waiting time, $p>0.05$ for effect of previous trial outcome and for the interaction, in either correct or error choices), (**Figure 3.S3A**). Contrary to previous results (Lak et al 2014) average waiting time significantly diminished from the beginning to the end of the session (two-way Anova, $p<0.001$ for the effect of odor contrast on

waiting time, $p < 0.05$ for effect of session period but $p > 0.05$ for the interaction, in either correct or error choices) (**Figure 3.S3B**), suggesting that average waiting time might be influenced by satiation. This difference in observations might come from the fact that while the number of trials per session is equivalent in both experiments rats might have experienced a higher amount of reward in our hands due to the reward manipulation of our task design.

Two factors predict the effect of reward magnitude manipulation on choice, accuracy and confidence report

To try to formalize in a computational model the different effects of our reward manipulation on choice behavior and confidence reporting we combined the mechanistic interpretation of metacognitive theories which implies a separation on the computations of choice and confidence estimation with the theoretical approach that confidence and choice can be derived by Bayesian and signal detection theory models (SDT) (Kepcs and Mainen, 2012). We took advantage of a model previously used to compute decisions and confidence in the 2AFC odor categorization task (Kepcs et al., 2008) and adapted it to generate predictions for our reward manipulation design.

In this model choices and confidence are computed by establishing a comparison between a perceived stimulus and the recalled category boundary (see Experimental Procedures). Stimulus and category boundary can assume a set of different values drawn from their specific distributions, which we assume to be Gaussian (**Figure 3.3A**). The stimulus distributions have the same variance but different means, in a odorA to odorB scale, depending on the mixture presented. The category boundary distribution has a given variance and a mean that is set around zero in trials where both options are equally rewarding (1A:1B). The first modification to the model was that we introduced a bias term, to represent the reward magnitude manipulation, that shifts the value sampled from the category boundary, positively (in 1A:3B blocks) or negatively (in 3A:1B blocks). In each trial a sample from stimulus distribution and a sample from the boundary distribution are compared

to derive choices **Figure 3.3B**). We adjusted the bias and category boundary distribution variance to better fit our behavioral results and we were able to generate psychometric curves (**Figure 3.3 top**), and choice behavior (**Figure 3.3 middle**) and performance (**Figure 3.3 bottom**) as a function of expected reward magnitude, which resembled our data.

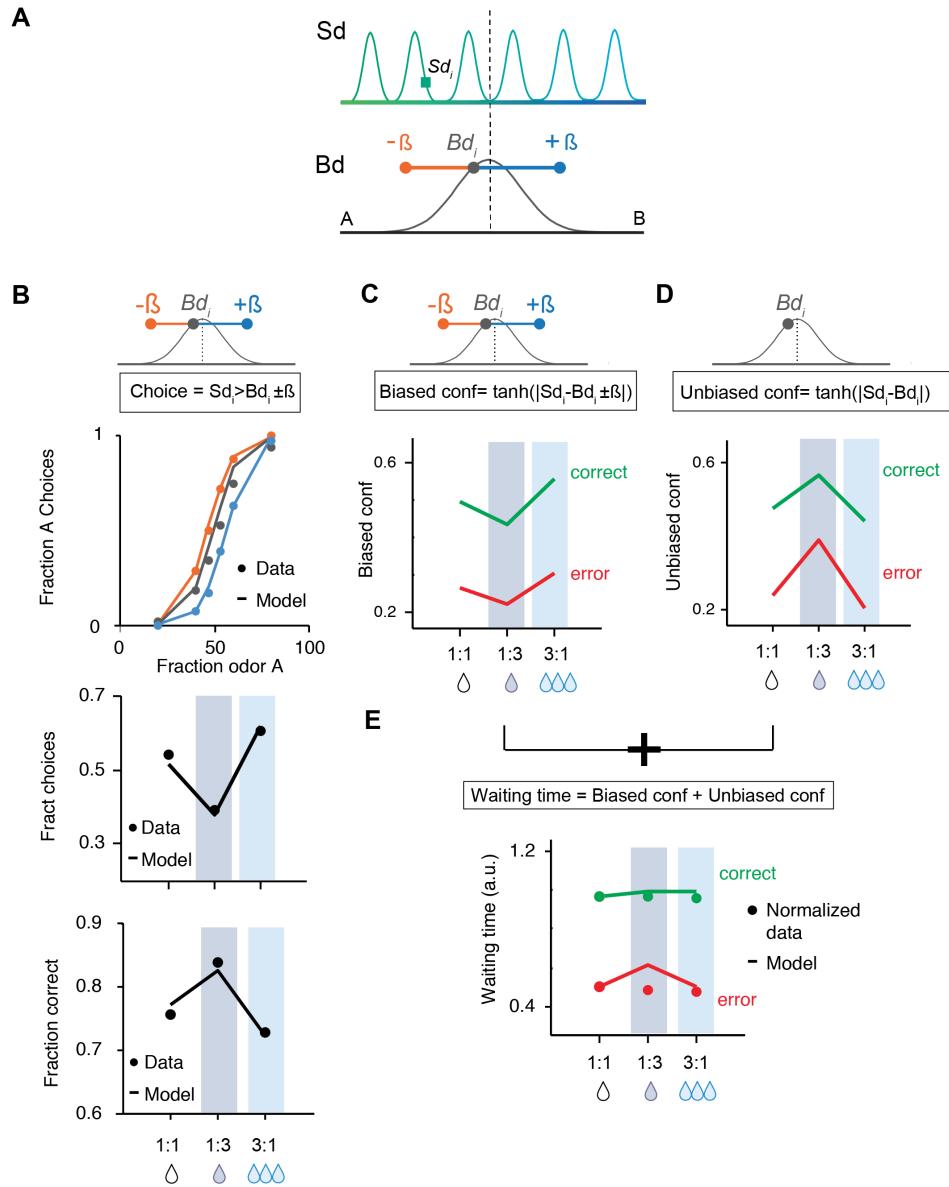


Figure 3.3 - Computation model for the effect of reward magnitude manipulation on choice and confidence report

(A) - Each odor mixture stimulus, as well as the memory for the category boundary, is encoded as a distribution of values – S_d and B_d . In each trial (i) a value is drawn from each distribution (S_{di} and B_{di}) and related to generate choice, confidence, and confidence report for that trial. To reflect the reward magnitude manipulation a bias β is introduced which shifts the boundary sample positively, for 3A:1B block trials (blue) or negatively, for 1A:3B block trials (orange).

(B) - This model can predict the psychometric curves previously obtained (*top*) as well as average choice behavior (*middle*) and accuracy (*bottom*).

Choices are calculated by comparing samples from S_d with samples from B_d . In 1A:1B blocks S_d is compared to the unbiased B_d . In reward manipulation blocks S_d samples are compared with samples $B_d \pm \beta$. (*top*) - Data points (filled circles) and model predictions (lines) are separated according to blocks trials, 3A:1B (orange), 1A:3B (blue) and 1A:1B (grey). (*middle, bottom*) - Data points and model predictions are separated according to value of choices- 1:1, 1:3 (purple) and 3:1 (blue).

(C) - Model predictions for biased confidence for correct (green) and error (trials) for 1:1, 1:3 and 3:1 choices. In each choice a biased confidence is calculated by the distance between S_d sample and B_d sample. In the reward manipulation blocks the B_d sample is biased by $\pm \beta$ and this is reflected in the biased confidence value obtained.

(D) - Model predictions for unbiased confidence for correct (green) and error (trials) for 1:1, 1:3 and 3:1 choices. In each choice an unbiased confidence, is calculated by the distance between S_d sample and the B_d sample, for all the trials.

(E) - Model prediction for waiting time for correct (green) and error (trials) for 1:1, 1:3 and 3:1 choices. In the model confidence report (waiting time) is computed for each choice by combining biased confidence and unbiased confidence values and the model predictions resemble the data.

In the previous version of the SDT model (Kepecs et al., 2008) each trial confidence is computed by the decision distance, the absolute difference between the stimulus and the category boundary samples. We introduced a second modification to the model, which consists in the notion that, in our reward manipulation task, there are two sources of confidence that the animals are using to generate the confidence report. For every decision the model has access to an internal representation of the quality of the decision, which takes into account the biased category boundary sample (**Figure 3.3C**). Confidence derived this way assumes a similar trait to what is observed for the reward manipulation effect on choice (**Figure 3.2A bottom**). In parallel the model keeps and uses the unbiased category boundary sample, to obtain a prediction of the reward probability given the stimulus, uncorrupted by the effect of reward bias, and uses it to compute an unbiased decision confidence (**Figure 3.3D**). Confidence derived this way assumes a similar trait to what is observed for the reward magnitude manipulation effect on accuracy (**Figure 3.2B bottom**). These two distinct sources of confidence can then be integrated to generate a confidence report (**Figure 3.3E**), and by doing so the model predicts very closely the waiting time data as a function of expected reward magnitude.

The model suggests a parallel computation of biased and unbiased decision confidence, which happens before a final waiting time can be generated. We have recorded electrophysiological activity of single units early on in the waiting period (**Figure 3.S4A**, Experimental

procedures) in two distinct brain regions, the ventral striatum (VS) and orbitofrontal cortex (OFC). The activity of a population of cells in these two distinct brain regions can predict outcome, with average firing rates being higher during error trials (*uncertainty*) (**Figure 3.S4B**). In VS activity of the *uncertainty* population, as a function of trial outcome and expected reward magnitude (**Figure 3.S4C**) resembled the biased decision confidence (**Figure 3.3C**) and in OFC (**Figure 3.S4D**) activity of the *uncertainty* population resembled the unbiased decision confidence (**Figure 3.3D**), suggesting a putative neuronal implementation of our model predictions in these two brain regions.

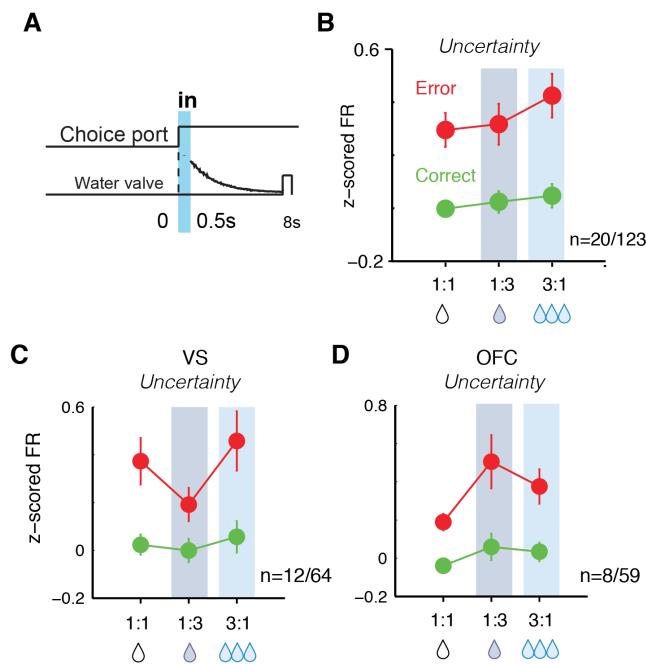


Figure 3.S4 – Uncertainty cells in VS and OFC and reward magnitude manipulation

(A) – Timing of outcome anticipation period. Neuronal activity was aligned to each trial choice port entry, signaled by the first break of photo-beams within each choice port, and analyzed for 500ms after choice port entry – light blue bar. During this period negative outcome predictive cells were identified in ventral striatum (VS) and orbitofrontal cortex (OFC) using ROC analysis (see Experimental Procedures).

(B) – Normalized firing rate of total uncertainty neurons, in correct (green) or error trials (red), for 1:1, 1:3 and 3:1 choices.

(C-D) – Normalized firing rate of (B) VS and (C) OFC uncertainty neurons, in correct (green) or error trials (red), for 1:1, 1:3 and 3:1 choices.

DISCUSSION

We have trained rats to perform a two-odor guided categorization task, which includes a post-wager decision confidence report. During each behavioural session they have experienced different reward schedules and we have analyzed how these affected rats behavior, in terms of their choices, accuracy and confidence report.

We have shown that when rats have to choose between two options leading to two different reward amounts their choices were biased by the

promise of a most valuable reward (**Figure 3.2A bottom**). The amount of bias was dependent of the amount of evidence available - higher bias was induced when odor contrast delivered was low (**Figure 3.1C**). The unbalanced reward schedule also led to differences in accuracy – rats were on average more accurate when choosing a lower valued option (**Figure 3.2B bottom**). But once animals committed to a choice we did not see a clear difference on the way they report their decision confidence - whether expecting a greater or a lesser amount of water rats waited on average the same amount of time (**Figure 3.2C bottom**), regardless of the amount of evidence available (**Figure 3.S1A**).

To explain this result we have hypothesized that waiting time was shaped by the interaction between two opposite sources: 1) a biased confidence (influenced by the expected amount of reward given the evidence) and 2) a unbiased confidence (which only takes into account the probability of being correct given the evidence). We managed to generate a model that predicted well our results by assuming that these two sources converge and give rise to a confidence report (**Figure 3.3**).

The model postulates that whenever an animal is asked to perform a two-alternative forced choice task the brain keeps track of a “true” confidence estimate, which takes into consideration the reliability and consistency of the accumulated perceptual evidence towards a decision, and which is not affected by the expected value of the alternatives. Given that the choice process does take into account the magnitude of expected reward – leading to biases in choice – this confidence estimate might not arise through the same mechanism that ultimately leads to a decision.

Our proposed model was based on principles from SDT that are widely used to formalize perceptual decisions in terms of computationally separable processes. One other set of models used to compute perceptual-guided decisions are models based on the accumulation of evidence towards a decision threshold. These models can be adapted to unfold decision confidence estimates by the difference between accumulated evidence for the two separated options (Kepecs et al., 2008). This framework might allow for different intuitions and give

insights into the underlying neural mechanisms of perceptual decision-making. To explain our results we can extrapolate the proposals of our model into the framework of evidence accumulation. We might hypothesize that in each trial the unbiased confidence estimate is computed early on in the decision process, before the difference in reward amount biases accumulation of evidence and the animal ends up choosing the most rewarded option. This unbiased confidence would have to be kept in working memory, in the end of the accumulation process the biased confidence would be computed and later on both confidence estimates would interact. However, this assumption would fail if we assume that reward biases influence the starting point of accumulation (Ding and Gold, 2013), or that, as previously shown (Wang et al., 2013), reward biases are being computed even before trial initiation.

A possible caveat of our study is that during behavioural training animals only experienced the reward manipulation schedule after having fully experienced and well performed the odor categorization. It might be possible that what we call unbiased confidence is not a trial-by-trial estimate but an estimate learnt due to previous outcome history. So, although in each trial a decision is made, taking into account the available evidence from that trial, and biased by expected reward, while the animal is anticipating reward the learnt outcome probability would affect the confidence report.

Whatever explanation holds, the results add-on to the notion that confidence reporting and the computation of perceptual decisions are distinct processes localized to different brain regions (Lak et al., 2014). They might also be somewhat concordant with the proposed view of a dual-route signal detection theory formalized for conscious versus non-conscious evidence accumulation and error detection (Charles et al., 2014; Del Cul et al., 2009)

We do not discard the view that perceptual-guided choices and confidence estimates are computed in the same brain region (Kiani and Shadlen, 2009) and relayed to other areas where the confidence-based

decisions take place (Insabato et al., 2010a). OFC is the best candidate region to play a central role in computing our confidence report (Kepecs et al., 2008; Lak et al., 2014). To better fit the ideas proposed by our model OFC should either locally compute the unbiased confidence, or keep track of the unbiased confidence computed somewhere else. Since OFC does not play a role in perceptual-guided decisions (Lak et al., 2014) (which were biased by our reward manipulation), it might be expected that the biased confidence is not locally computed in the OFC but relayed to here from some other region.

Related to this we have found a population of neurons in the OFC whose activity pattern was qualitatively similar to the unbiased confidence prediction, just after choice port entry (**Figure 3.S4D**). In parallel, we have found a population of neurons in the VS that had an activity pattern qualitatively similar to biased confidence, during the same trial epoch (**Fig 3.S4C**). Merging both populations together held an activity pattern qualitatively similar to our waiting time result. Given this one view on the neural mechanism of computing the confidence report would be that both types of confidence signals could actually be forwarded, not to the OFC, but to one downstream brain region responsible for the leaving decision, such as the pre-motor cortex (Murakami et al., 2014). Nevertheless, we cannot rule out that we simply missed the small OFC neuronal population previously reported to correlate with both expected value and reward uncertainty (O'Neill et al., 2010). If we assume that unbiased confidence is related to perceptual uncertainty (and that perceptual uncertainty in our task is somewhat correlated with reward uncertainty) this population could be a good candidate to have activity correlated to the convergence of the two proposed opposite confidence signals. One other putative neural implementation of our mechanism would be the circuit proposed in value-based decisions in humans (De Martino et al., 2013), in where subjective confidence (unbiased confidence) and difference in value (bias) are computed differently in the ventromedial prefrontal cortex (human analogous to rodent OFC) and accessed by the rostral-lateral prefrontal cortex to enable a confidence report.

Our results also revealed some interesting dynamic features of the role of reward and confidence on animals' behavior. It was previously known that in a two-odor categorization task the motivation to start a trial is higher when more net value is expected, and choices are biased towards the most rewarding choice (Wang et al., 2013). Further on the reward bias not only influences choice probability but also keeps influencing animals' behavior during movement time (**Figure 3.S2B**). Animals move faster when going towards a more valuable option but then do not wait longer for reward. In our paradigm the probability of making a correct decision is not clearly reflected in the behavior before a choice had been committed (Uchida and Mainen, 2003; Zariwala et al., 2013) but it is robustly reflected after that, in the waiting time of every tested animal (**Figure 3.S2A**). As previously reported (Busse et al., 2011) choices were influenced by the interaction between available evidence and outcome previously experienced; the same is true for the amount of reward obtained in the previous trial (data not shown). But previous trial reward amount does not affect waiting time (**Figure 3.S3A**). This adds to the notion that waiting time is a trial-by-trial confidence report that can provide resolution to the measure of quality of a decision just made, without being affected by previous experiences.

In conclusion, we have explored the interplay between value and confidence ($P(\text{correct}|\text{evidence})$) in the behavioural manifestation of a confidence report. We have observed that although choices were biased by expected reward magnitude our confidence report did not seem to be affected by it. Nevertheless it did not perfectly predict decision accuracy. We have suggested that value and confidence were opposite forces which combined gave rise to the observed confidence report. Or, in other view, introducing a choice bias by reward manipulation is one factor that can alter the calibration of confidence in perceptual judgments (Johnson and Fowler, 2011).

This work has implications on the way we think about the effects of reward in perceptual decision-making. When allured by bigger rewards our choices are clearly biased. These biases change our performance but,

while unaware of the outcome, we might not be able to optimally adapt our behavior to reflect the quality of our decision - our confidence report is badly calibrated with accuracy. When waiting to know the outcome of a decision, if we have a hold on how much our decisions were biased by reward we might be able to extract a better notion of decision quality out of our confidence report. We could then change our behavior accordingly, preventing the spread of a bad choice or persevere on waiting in order to profit from a “truly” good decision.

EXPERIMENTAL PROCEDURES – II & III

Subjects

A total of 6 male adult Long-Evans rats were used for the experiments. Rats were motivated by water restriction and had unlimited access to food. All procedures involving animals were carried out in accordance with European Union Directive 86/609/EEC and approved by Direcção-Geral de Veterinária.

Behavior

Behavioral apparatus

The apparatus has been described previously (Uchida and Mainen, 2003). The behavioral box contains a panel of three ports: the central port for odor delivery ('odor port'), and two ports on each side ('choice ports') for water delivery. 1 infrared photo-beam detector positioned on the inner wall of each port detected entry and exit from the ports. A custom-designed computer controlled airflow dilution olfactometer was used for odor delivery. Saturated odor vapor was produced by flowing clean air through disposable syringe filters (glass microfiber, pore size = 2.7 µm, #6823-1327, Whatman) that were loaded with 20 µl of 1/10 mineral oil dilutions of liquid odorants. Filters were inserted into a PEEK manifold of all valves and controllers. Odor streams were mixed at the manifold directly before the odor port with a clean air stream ('carrier') to produce a total flow rate of 1 L/min (1:20 dilution). The behavioral box is controlled using custom-made Matlab (Mathworks) software communicating with a real-time state-machine controlling the olfactometer and data acquisition hardware card (National Instruments) running on a Real-time Linux operating system.

Behavioral task with confidence report

Rats self-initiated each experimental trial by introducing their snout into the central port where odor was delivered. After a variable delay, drawn

from a uniform random distribution of 0.2–0.5 s, in 95% of the trials, a binary mixture of two pure odorants, S(+)-2-octanol and R(-)-2-octanol (Sigma), was delivered. Per block of 40 trials rats experienced the following 6 concentration ratios, arranged pseudo randomly: 80:20 – 3/40 of the trials, 60:40 – 12/40 of the trials, 53:47- 4/40 of the trials, 47:53 - 4/40 of the trials, 40:60 – 12/40 of the trials, and 20:80 - 3/40 of the trials. Also in the same odor block a binary mixture of two pure odorants (p-cymene and anisole (Sigma)) were delivered in 2/40 trials but further excluded from analysis. This odor block delivery scheme was used to ensure that rats experienced at least twice the full diversity of stimuli in each of the three reward manipulation blocks.

After a variable odor sampling time up to 1 s, rats responded by withdrawing from the central port, which terminated the delivery of odor, and moved to the left or right choice port. Choices were rewarded according to the dominant component of the mixture, that is, at the left port for mixtures A/B > 50/50 and at the right port for A/B < 50/50. A variable reward delay period after entry into the choice port was introduced. For correct choices, reward was delivered between at least 0.5 s after entry into the choice port and up to 8 s. The reward delay was drawn from an exponential distribution with decay constant equal to 1.5 resulting in a relatively constant level of reward expectancy over a range of delays (i.e. flat hazard rate). In a small fraction of correct choices (10% - 15% of correct trials) distributed pseudo randomly throughout the behavioral session the reward was never delivered to the rat (catch trials). These trials were designed to never occur consecutively. As the animal spends time to consume the water in the rewarded correct trials, we used the catch trials to measure WT for correct choices. After one second from water delivery or from choice port exit in unrewarded trials rats could initiate a new trial. The duration of each day behavioral session was not strictly defined, and session duration was thus less predictable for the animal.

Reward manipulation

In each session rats experienced three different conditions/blocks of reward delivery in which the ratio of reward magnitudes took one of three values - 1A:1B, 3A:1B, or 1A:3B (where A and B are the choice ports, and 1 and 3 are the amounts of water delivered, 24 μ l and 72 μ l respectively). Each session started with 140 trials of the 1A:1B block, followed by 120 trials of either 3A:1B or 1A:3B blocks. For the next 120 trials the larger reward changed spatial location. After this, and until the end of the session the reward block switched back to 1A:1B. The order of the first reward manipulation block (3A:1B or 1A:3B) was counterbalanced from session to session.

Behavioral training

In order to be able to perform the perceptual categorization task described above, rats went through a multistep training procedure typically lasting 6-8 weeks. Initial steps of training consisted of imperative trials in which animal was supposed to poke into the odor port and collect the water reward subsequently from either of the two choice ports. Choice trials were gradually introduced as the training proceeds; starting with easy decisions between pure odor stimuli and advancing gradually to more difficult trials in which odor mixtures close to decision boundary (lower odor mixture contrast) were introduced to the animal. During initial training sessions reward followed the correct choices immediately. After animals reached psychometric performance (>95% for pure odors) delayed reward delivery was gradually introduced in successive behavioral sessions, followed by the introduction of reward omission in the trials. At this time pure odors were no longer introduced and rats experienced only 60%, 20% and 6% odor mixture contrasts. When the confidence report was stable, rats started experiencing the blocked-wise manipulation of reward magnitude. Behavioral data was collected after animals had experienced 4 alternated sessions of reward manipulation for a total of 213 sessions (min. 26, max. 40 sessions per

rat) – Chapter III. After this period rats were prepared for surgery – Chapter II.

Analysis of behavioral data – Chapter II

During the collection of electrophysiological data, animals performed on average 483 trials per recording session, and data was collected during a total of 28 sessions (5 to 7 sessions per rat).

In order to estimate rats' decision accuracy as a function of waiting time, we assumed that waiting times for correctly performed catch trials (which were pseudo randomly distributed) were a good representative for the distribution of all correctly performed trials. Therefore the catch trials waiting time data were expanded to all correct trials (taking into account the odor stimulus identity). Each session trials were then grouped according to the waiting time distribution into 7 equally spaced bins and accuracy was calculated for each bin.

Each individual rat data was an average over the median values for each behavioral session in the case of waiting time and movement time, or an average of session's fraction choices (for accuracy, psychometric function). Error bars in figures indicate standard error of mean (s.e.m) across sessions for individual animals or across rats for the population data.

Analysis of behavioral data – Chapter III

In total, we've analyzed 106631 trials from 6 rats from a total of 205 sessions (8 sessions were excluded because rats did not experience the three different reward blocks). Rat's performed an average of 391 trials per session (min 200 trials, max 747 trials), where on average 179 trials were evenly rewarded and on average 203 trials were unevenly rewarded.

For statistical analysis across rats, averaged data for each rat was used. Each rat data was an average over the median values for each behavioral session in the case of waiting time and movement time, or an average of session's fraction choices (for accuracy, psychometric function and

choice bias). Error bars in figures indicate standard error of mean (s.e.m) across sessions for individual animals or across rats for the population data. To compute statistical significance we used nonparametric Wilcoxon signed rank test for single comparisons and two-way ANOVA, for multiple comparisons.

In figure 2.2 top two rows, sliding averages (of fraction choices, accuracy or waiting time) with 50 trial windows were computed for each session of rat Go10. Averages and s.e.m over the sessions were then calculated.

Only waiting times from error trials or catch trials were used. The distribution of waiting times was somewhat bimodal, especially for error trials (Fig 3S5). For all the analysis we have excluded waiting times shorter than 1.2s and longer than 15s. Due to the bimodality of the distribution we have considered that trials in which leaving decisions were done before 1.2s of entering the choice port (0.7s of waiting, since water could only be delivered after a 0.5s delay) were trials where choices had been qualitatively different from the rest of the dataset (in 9% of error trials, 1.5% of catch trials, for all odor contrasts). These error trials might be driven by exploratory decisions. On the other hand trials longer than 15s were considered to be behavioral setup acquisition mistakes, most likely related to maintenance of photo beam disruption due to debris (0.2% of error trials and 0.3% of catch trials). Nevertheless, including the excluded trials showed comparable results (data not shown).

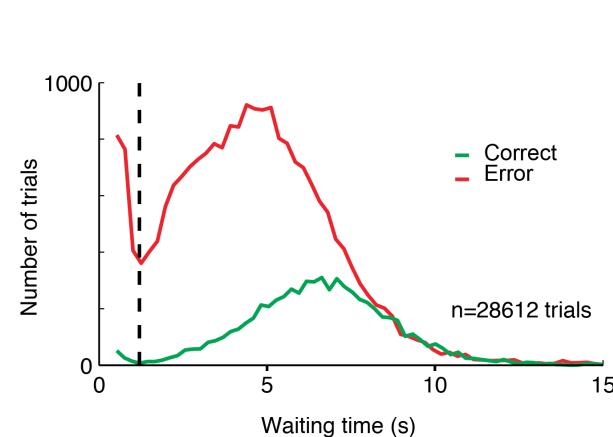


Figure 3.SS5 – Distribution of waiting times

(A) - Histogram of waiting times in the behavioral dataset. Waiting times from a total of 6565 correct catch (green), and 22047 error trials (red) were plotted. For analysis waiting times shorter than 1.2s (dashed line) and longer than 15s were excluded – 9% of error trials and 2% of catch trials.

The effect of previous reward experience on waiting time was tested by conditioning waiting time on trial t to the amount of reward obtained in trial $t-1$. For this we considered that animals could have experienced three different outcomes: 24 μ l of water, in $t-1$ 1:1 or 1:3 choices, 72 μ l of water in $t-1$ 3:1 choices or no water, in $t-1$ error or catch trials.

To test the effect of within-session satiation on waiting time we divided each session trials, (taking into account the odor stimulus identity) into 30% percentiles and compared waiting time in the trials within the 1st 30% percentile (session beginning) with waiting time within the 3rd 30% percentile (session end).

Neural recordings

For the electrophysiology experiments, each rat was implanted with a drive (Island motion corporation, Tappan, NY) with a double cannula, allowing for two distinct parallel recording sites. Each cannula directed 10 to 12 movable tetrodes targeting the left hemisphere OFC (3.5mm anterior, 2.5mm lateral) or the left hemisphere VS (1.25 mm anterior, 2mm lateral). In the surgery day the implanted cannulas remained at the surface of the craniotomy cavity, and the individual tetrodes were moved until reaching a depth closer to the final target OFC (5.44mm), VS (8mm).

Individual tetrodes consisted of four twisted polyimide-coated nichrome wires (H.P. Reid, Inc.; single-wire diameter 12.5 μ m) gold-plated to 0.2-0.5 M Ω impedance at 1 kHz. For better stability and targeting each implanted tetrode was covered by a 0.0055" OD polyimide tube (MinVasive Components), with the tetrode tip extending 1-1.5mm from the tube tip. Tetrodes were coated with DiI (Molecular Probes) to visualize the tetrode tracks in a histological examination.

Tetrode depths were adjusted before or after each recording session in order to sample an independent population of neurons across sessions. The locations of tetrode tips during each recording session were

estimated based on their depth and histological examination based on electrolytic lesions and the visible tetrode tracks.

Electrical signals were amplified and recorded using the Cerberus data acquisition system (Blackrock Microsystems). Neural and behavioral data were synchronized by acquiring time-stamps from the behavioral system along with the electrophysiological signals.

Histology

In order to verify the ultimate location of the tetrodes, electrolytic lesions were produced after the final recording session ($15\mu\text{A}$ of cathodal current, 10s). Rats were then deeply anesthetized with pentobarbital and perfused transcardially with 4% paraformaldehyde. The brain was removed from a skull, stored in 4% paraformaldehyde, sectioned at 50 μm . Every other slice was stained with Cresyl violet solution with a standard Nissl staining protocol to observe sites of electrolytic lesions. Other slices were prepared for fluorescent observation to examine the fluorescent tracks made by DiI-coated tetrodes.

Neural data analysis

Multiple single units were isolated offline by manually clustering spike features derived from the waveforms of recorded putative units using MCLUST software (A.D. Redish). From the isolated population of cells 69 were excluded for further analysis for not being stably recorded across the session, due to cell drift or interruption of the signal acquisition while the rat was still behaving. A remaining 22 cells were excluded for being located outside of the targeted areas, in the piriform cortex. In the end of the clustering procedure a total of 123 individual neurons were considered of good quality to be used for further analysis. Data analysis was performed using Matlab.

We focused our analysis mainly in the ‘reward anticipation period’ – activity aligned at choice port entry and measured for a window of 500ms after choice port in. This excluded spikes that occurred during or after water valve actuation on correct trials; on error trials or catch trials,

no feedback was present. Two other epochs were analyzed to compare their activity with that of the reward anticipation period. For analysis of the ‘before waiting period’ activity was aligned at choice port entry and measured for a window of 500ms before choice port in. During this time rats were most likely moving from odor port to choice port. For analysis of the ‘end of waiting period’ activity was aligned at choice port exit and measured for a window of 500ms before choice port out. At this time rats were most likely experiencing water delivery, for correct trials, or aware of reward absence, for error and catch trials.

For visualization purposes in raster plots trials were sorted according to movement time in ascendant order and the peri-event-time-histograms (PETHs) were smoothed with a Gaussian filter (s.d.= 25 ms). Activity is presented as mean firing rate and s.e.m. across trials, for individual units, and across cells, for population data.

To determine how well neural activity predicted the upcoming outcome (correct-reward/error-no reward), we used receiver operating characteristics (ROC) analysis to calculate an outcome relative preference index (OP) that measures how well an ideal observer can predict the outcome from the knowledge of the firing rate from trial to trial (Feierstein et. al. 2006, Felsen and Mainen, 2008, Kepecs et al. 2008). This index varies from -1 to +1 with the sign denoting whether a neuron fires more for correct (+) or error (-) decisions:

$$OP = 2 \times (ROC_{area} - 0.5); ROC_{area} = \int_0^{\infty} P(f_{correct} = f)P(f_{error} < f)df$$

where $f_{correct}$ and f_{error} refer to the distribution of firing rates during the analysis period in correct and error trials respectively. Statistical significance was evaluated using a permutation test, where trial order was pseudo-randomly shuffled 1000 times to yield a p value. Neurons with $p < 0.05$ were denominated *uncertainty* cells (negative OP) or *confidence* cells (positive OP).

To relate the activity of uncertainty population of cells, or confidence population of cells, with accuracy and waiting time we have grouped

each unit's trials according to it's z-scored firing rate, binned into 7 equally spaced bins. We then calculated accuracy, or average z-scored waiting time, for each group of trials. To extract statistical significance we have calculated the correlation coefficient between the averaged z-scored firing rates and accuracy, or averaged z-scored waiting time, and computed a t-statistics based p-value for the hypothesis of null-correlation.

To better visualize the uncertainty or confidence population activity, dependent of waiting time, we have grouped each unit's trials according to waiting time into 5 groups with equal amount of trials (waiting time percentile ranks). We have generated PETHs of z-scored firing rates for each group of each cell, and then averaged the PETHs of the population of neurons, for each waiting time percentile separately.

We have also analyzed each individual neuron's firing rate during each one of the trial periods and it's relation with waiting time. For this we have done a linear regression analysis and explored if each trial's waiting time (WT_i) could be a good linear predictor of each neuron's average trial firing rate during the different analyzed periods:

$$Rate_i = \beta_0 + \beta_{wt} \times WT_i$$

where β_0 is a constant term and β_{wt} is the regression coefficient for waiting time as a predictor. We used F-test to select for neurons ($P < 0.05$). The F-test tests whether a proposed regression model as a whole fit the data significantly better than a simpler model (firing rate = β_0).

Using the activity analyzed in each one of the three epochs, we have correlated individual (OFC or VS, separately) neuron's waiting time linear regression coefficients with the corresponding outcome relative preference indices and computed a t-statistics based p-value for the hypothesis of null-correlation.

Confidence report and value: computational model

Our computational model is an adaptation of a previously developed model for two-alternative decisions that uses a signal detection theory framework. In this model each choice and its associated confidence could be estimated by comparing a sampled stimulus and a decision boundary (Kepecs et. al. 2008, Lak et. al. 2014).

We have modeled the internal representation of a presented stimulus, in each trial i , as the log ratio of the presented odor mixture with additive Gaussian noise:

$$S_i = \log \frac{[A_i]}{[B_i]} + \sigma_s$$

where $\sigma_s = 0.2$, for all the different stimulus. The probability of stimulus occurrence was matched to what was experimentally set (Experimental Procedures II & III- Behavioral task with confidence report). The boundary distribution is modeled also with additive Gaussian noise $\sigma_b = 0.4$, with mean 0 (50/50% mixture).

The blocked-wise reward manipulation was modeled by adding a bias term, β_{block} , to the decision boundary sample, which varied according to the reward manipulation block in where the decisions were made. We have set β_{block} values to better fit the choice bias occurring during the different reward manipulation blocks: $\beta_{3A:1B} = -0.15$, $\beta_{1A:1B} = -0.05$ and $\beta_{1A:3B} = 0.25$. In each trial i a sample from the stimulus distribution (Sd_i) and a sample from the boundary distribution (Bd_i) were randomly drawn. To obtain the results (Fig 3) the model made a total of 100 000 trials.

Choices were computed by comparing Sd_i and Bd_i :

$$choice_i = \left\{ \begin{array}{l} sideA \mid Sd_i > Bd_i + \beta_{block}; \\ sideB \mid Sd_i \leq Bd_i + \beta_{block} \end{array} \right\}$$

In each trial two types of confidence estimates were generated. One confidence estimate took into account the different reward magnitude manipulations, reflected by the β_{block} term, which induced the choice bias:

$$Biased\ conf_i = \text{tahn}(|Sd_i - Bd_i + \beta_{block}|)$$

where *tahn* is a scaled logistic function previously used to calibrate confidence to the outcome probabilities (Kepecs et. al. 2008). The second confidence estimate is unaffected by the reward manipulation. This unbiased confidence can be thought of as a more robust representation of the outcome probability expected given the presented stimulus.

$$Unbiased\ conf_i = \text{tahn}(|Sd_i - Bd_i|)$$

To obtain each trial confidence report the model took into account the two types of confidence and simply added them together.

$$WT_i = Biased\ conf_i + Unbiased\ conf_i$$

The specific properties of the timing element in our confidence report (waiting time) were not considered but could be possibly better formalized using other types of models that take into account variables such as the relation between hazard rate and waiting cost functions (Lak et al., 2014).

GENERAL DISCUSSION

Waiting time as a proxy for confidence

We have approached the cognitive-loaded concept of confidence by studying perceptual decision-making and confidence-based behaviors in behaving rodents. Additionally, we have developed normative models to infer optimality on a time waging behavior that depends on confidence and investigated the role of value in a confidence report. Also, we have causally linked the OFC of rats to the report of decision confidence. Moreover we have analyzed individual neurons activity correlated with decision confidence and confidence report, both in the OFC and in the VS.

In order to better understand the neural and behavioural mechanisms of perceptual decision confidence, we have trained rodents to perform a classic odor guided two alternative forced choice task (Uchida and Mainen, 2003). We have adapted this task such that, after a decision is made, rodents need to optimally wager on the time they are willing to wait for a reward. Waiting time was a proxy for decision confidence and is likely to be a better report of confidence, when compared to a binary wager or a opt-out task (Kepecs and Mainen, 2012). Mostly given that for each trial it is possible to get a hold on both decision and confidence estimates.

Unpublished data from our lab (Venturini et al., 2014) shows that humans engaged on a visual-guided decision task can also use decision confidence to wager with time, and that the waiting time correlates well with a scalar confidence report. This waiting time paradigm could be well fit to study human decision confidence and metacognition, providing a better comparison between human and non-human animal data, given that it does not imply any semantic report.

We have used two different models to describe how could a waiting time-based confidence report be generated from a perceptual decision. In the first chapter of this dissertation we have shown that we could quantitatively establish the relation between optimal waiting time and reward probability, which in perceptual decisions can be estimated based

on the confidence associated with a decision. This could well predict animals' waiting time data. Waiting time was also dependent on the estimated reward delivery, an indication that reinforcement-related factors also influence the post-decision wager.

A possible separation between perceptual decisions and confidence estimation

In order to evaluate the role of OFC in decision confidence reports we have pharmacologically inactivated this brain region in rats performing the waiting time task. Waiting time dependency on decision confidence was disrupted. It was not stated, but might be suggested, that after inactivation of orbitofrontal cortex (OFC) the rat's estimation of optimal waiting time was mostly based on the estimation of reward delivery time, which should be independent of stimulus delivered. To better understand the computation of optimal waiting time, it was left to model how other variables related to reward such as reward magnitude, influence subjective estimation of reward delivery time or the opportunity cost of waiting. The results obtained in the third chapter of this dissertation could be used to test predictions obtained from such a model.

To further explore the relation between waiting time and reward, we have manipulated reward magnitude expectation. In a block-wise design we have increased the value of one of the choices and analyzed the effect of this on animals' behavior. Animals did not wait differently when expecting more or less reward, but were biased towards the choice offering larger amount of reward, and were more accurate whenever choosing the least rewarding option.

We have used a SDT-based framework to explain this result, and introduced the notion of two opposite factors which influence waiting time - a unbiased factor, related to the reward probability only indicated by the stimulus, and a biased factor, which includes the both reward probability and reward magnitude. The interaction of these two factors

establishes a confidence report. The model could well predict choice, accuracy and waiting time behavioural data.

This two-factor model has limitations, which arise mainly from its simplicity. For instance, unlike the model presented in the first chapter of this dissertation, it does not address the dynamics necessary to estimate timing of leaving decision. Also the way the two factors are combined to generate the confidence report can be seen as somewhat artificial. But we would like to point that the relevancy of this model comes not from its sophistication but from what it proposes: a dual-route for the computation of decision confidence, which enforces the view of a possible separation between perceptual decisions and confidence estimation.

A dual-route in perceptual decisions would assume the existence of one system responsible for categorizations, and a separate system responsible for supervision of these categorizations. This separation has been proposed by others, in the context of conscious vs. unconscious strategic judgments (Bechara et al., 1997; Charles et al., 2014; Cul et al., 2009; Dehaene et al., 2006), reward-based decision making (Fleming and Dolan, 2012; De Martino et al., 2013), signal detectability (Galvin et al., 2003) or perceptual decision making (Insabato et al., 2010a; Zylberberg et al., 2012).

The separation between perceptual decisions and confidence estimation was further enforced by the results from OFC inactivation, which disrupted waiting time, without affecting the odor-guided decision. Given this result we suggest that OFC can be regarded as a hub for centralization of confidence levels, alongside other reward-related variables. OFC could be implementing the necessary operations to compute waiting time proposed in the two-factor model. It's waiting time correlated activity could serve as evidence for leaving decisions, computed in secondary motor cortex (Murakami et al., 2014).

Future work should try to combine both models presented in this dissertation, to better establish the theoretical framework which can not

only accurately predict the use of decision confidence estimates to derive optimal waiting but also to mechanistically determine the concrete role of OFC in metacognitive judgments.

A question subsides when looking at the results from the first and third chapters of this dissertation - what would be the effect on waiting time when inactivating OFC during the reward magnitude manipulation task? The block design used in this task makes its so that the amount of reward offered in each port is stable and predictable across the block. More so, these amounts are independent of the stimulus delivered. If OFC role is to compute outcome expectations “on the fly”, regardless of previous value experience, inactivating OFC could result in the disruption of the relation between decision confidence and waiting time, without affecting the relation between reward magnitude and waiting time. In this case animals would wait longer when expecting higher rewards. Instead inactivation of OFC could disrupt the dependency of waiting time with both reward magnitude and reward probability. Waiting time would then reflect other hidden variables, computed outside OFC.

OFC and VS in estimating decision confidence

Apart from OFC (Kepcs et al., 2008), neuronal signals related to decision confidence have been observed previously in other brain regions (Kiani and Shadlen, 2009; Komura et al., 2013; Middlebrooks and Sommer, 2012). We have included the basal ganglia in the plethora of regions that can compute decision confidence. We have found neuronal populations in the ventral striatum (VS) whose activity correlates with decision confidence (or uncertainty), and also with waiting time.

VS was previously implicated in inferring value information necessary to drive reward-based decisions (Bissonette et al., 2013; Cromwell and Schultz, 2003; Fitzgerald et al., 2014; Haber and Knutson, 2010; Ito and Doya, 2009; van der Meer and Redish, 2011; Meer et al., 2009; Roesch et al., 2009) and is suggested to play the role of critic, in model-based reinforcement learning, monitoring decisions and computing outcome expectations (Ito and Doya, 2011; Mannella et al., 2013; van der Meer

and 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). The results presented here suggest that VS also has access to measures of decision uncertainty to play that role.

Moreover, OFC and VS might work together, as a cortico-striatal circuit for implementing optimal behavioural strategies that take into account decision confidence estimates. To further explore the role of VS in the computation of confidence estimates and the behavioural use of these estimates it would be useful to manipulate neuronal activity in this region. This could be done with the same inactivation strategy used for OFC. Or, preferably, it could be done using optogenetics, a strategy that can give insights about within-trial temporal dynamics and striatal-circuit specificity. A well-controlled manipulation would allow for a better mechanistic description of the roles of both OFC and VS in computing decision confidence estimates, reward-based decisions and behavioural adaptation.

Future work: where is noise in the circuitry?

We suggest that OFC and VS integrate distinct sources of information, relayed from other areas, to provide outcome predictions based on confidence monitoring processes. But the question of where does the uncertainty, or noise, arises in the olfactory-guided categorization still remains to be answered. Generally, it can be assumed that uncertainty might arise from noise at the sensory representation of task-related stimuli or from noise at the regions responsible for associating these representations with a categorical decision, and these two sources can be computed in separate brain regions (eg. Vilares, Howard, Fernandes, Gottfried, & Kording, 2012). Previous evidence suggests that in the odor categorization task noise does not arise at the level of sensory discrimination (Miura et al., 2012b), but from uncertainty about the precise category boundary, which has to be learnt by the subject through trial-by-trial reinforcement (Zariwala et al., 2013). This category boundary could be represented in a neuronal layer that receives

projections from both a sensory layer representing stimulus information (eg. the piriform cortex) and projections from reinforcement signals important to update value of actions (eg. ventral tegmental area).

Results from a visual guided reward-based decision task performed by monkeys have implied that the oculomotor caudate (primate equivalent of dorsal striatum) contributes to the formation of the decision variable early in the decision process, biasing actions based on value properties (eg. Lauwereyns, Watanabe, Coe, & Hikosaka, 2002). Moreover, corticostriatal circuitry, involving projection neurons from sensory cortex to dorsal striatum have been implicated in biasing actions according to sensory stimuli, at least in auditory-guided decision making (Znamenskiy and Zador, 2013). Given this we can predict that in the odor categorization task, the category boundary could be represented in a corticostriatal circuitry involving the piriform cortex and a specialized olfactory striatum – the olfactory tubercle. The olfactory tubercle is an extension of the ventral striatum that receives projections from the piriform cortex and VTA and may use reward-based learning rules to encode odor valence (reviewed in Giessel & Datta, 2014; Ikemoto, 2007). Hence, probing into piriform-to-tubercle projection neurons could result in identifying the source of noise that limits performance in odor-guided categorizations. Alas, although plausible, this might be a daunting hypothesis to pursue. Mainly due to the fact that tubercle neurons, which receive projections from piriform cortex, are situated deep in the most ventral part of the rat's forebrain. Adding to this, targeting the appropriate projections should be difficult since the representation of odors in piriform cortex, and maybe also in tubercle, are widely dispersed and do not follow a topographical organization (Stettler and Axel, 2009).

Concluding Remarks

In the work presented in this dissertation we have explored the behavioural and neural mechanisms of decision confidence. Confidence judgments, self-assessments about the quality of a subject's decisions, beliefs or knowledge, are considered a central example of metacognition.

Failures in metacognition can result in lack of insight or unawareness of illness in many psychiatric disorders, such as schizophrenia, addiction, bipolar disorders and other psychoses (reviewed in David, Bedford, Wiffen, & Gilleen, 2012). Misevaluation of decisions and failure to adapt behaviors is also a feature of patients with frontal-lobe lesions. (Bechara et al., 1997; Dehaene & Changeux, 2011; a Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009). Understanding the neural mechanisms underlying decision confidence and confidence-based judgments can shed lights on the basis of the covert neurobiological processes disrupted in mental illness.

Confidence estimates are important in everyday life situations, not only for subjective evaluation of individual decisions, but also in the context of social interactions. The way we report decision confidence to others can have social implications, and is strategically important in the context of competition for limited resources (Johnson and Fowler, 2011). Also, shared confidence evaluation seems to increase performance in decisions, increases the accuracy of our own confidence report and is important in collaboration efforts involving outcome or perceptual uncertainty (reviewed in Frith, 2012). Moreover, the importance of metacognition in education is widely acknowledged, and some efforts have been proposed in order to enhance metacognitive abilities, with implications in learning linked to student's conceptions about their own knowledge (reviewed in Thomas, 2002).

Undisputedly, the study of the neurobiology mechanisms of decision making in animal models can expose general principles of cognitive function (Shadlen and Kiani, 2013). The study of cognitive function is fundamental in the attempt to decipher nervous system function , allowing for approaches which encompass from molecules in individual cells to complex behaviors (Laurent and Fregnac, 2014; Mainen and Pouget, 2014). When neuroscience studies animal behaving it infers brain function from correlated neural activity or causal disruptions of behavioural features. As neuroscientists we grasp multiple elements of knowledge but are however not aware of the entire picture. The same is

true when we look at parsed circuits, populations of neurons or single cells and its molecular features. Whatever levels of research we focus on, let us not forget that the whole is greater than the sum of its parts, as the wise man said. In the end, it is the sum of our neurons and synapses, circuits and systems that allows us not only to behold the swimming Euglena struggling in the pond, but also to wonder about our own thoughts, beliefs and decisions.

REFERENCES

- Abbott, A. (2010). THE RAT PACK. *Nature* *465*, 282–283.
- Balleine, B.W. (2011). Sensation, Incentive Learning, and the Motivational Control of Goal-Directed Action. in *Neurobiology of Sensation and Reward* (Boca Raton(FL): CRC Press).
- Balleine, B.W., and Dickinson, A. (1998). The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. *26*, 46–59.
- Balleine, B.W., and O'Doherty, J.P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* *35*, 48–69.
- Balleine, B.W., Leung, B.K., and Ostlund, S.B. (2011). The orbitofrontal cortex, predicted value, and choice. *Ann. N. Y. Acad. Sci.* *1239*, 43–50.
- Baranski, J. V, and Petrusic, W.M. (1994). The calibration and resolution of confidence in perceptual judgments. *Percept. Psychophys.* *55*, 412–428.
- Barto, A., and Sutton, R. (1998). Reinforcement Learning an Introduction, in *Adaptive Computation and Machine Learning*. (MIT Press).
- Bechara, a, Damasio, H., Tranel, D., and Damasio, a R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science* *275*, 1293–1295.
- Beck, J.M., Ma, W.J., Kiani, R., Hanks, T., Churchland, A.K., Roitman, J., Shadlen, M.N., Latham, P.E., and Pouget, A. (2008). Probabilistic Population Codes for Bayesian Decision Making. *Neuron* *60*, 1142–1152.
- Berlin, H.A., Rolls, E.T., and Kischka, U. (2004). Impulsivity , time perception , emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. *Brain* *127*, 1108–1126.
- Bissonette, G.B., Burton, A.C., Gentry, R.N., Goldstein, B.L., Hearn, T.N., Barnett, B.R., Kashtelyan, V., and Roesch, M.R. (2013). Separate populations of neurons in ventral striatum encode value and motivation. *PLoS One* *8*, e64673.
- Botvinick, M.M., Niv, Y., and Barto, A.C. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* *113*, 262–280.
- Britten, K.H., Shadlen, M.N., Newsome, W.T., and Movshon, J.A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* *12*, 4745–4765.
- Britten, K.H., Shadlen, M.N., Newsome, W.T., and Movshon, J.A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.* *10*, 1157–1169.

- Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science* *340*, 95–98.
- Burke, K.A., Franz, T.M., Miller, D.N., and Schoenbaum, G. (2008). The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards. *Nature* *454*, 340–345.
- Burke, K.A., Takahashi, Y.K., Correll, J., Brown, P.L., and Schoenbaum, G. (2009). Orbitofrontal inactivation impairs reversal of Pavlovian learning by interfering with “disinhibition” of responding for previously unrewarded cues. *Eur. J. Neurosci.* *30*, 1941–1946.
- Busse, L., Ayaz, A., Dhruv, N.T., Katzner, S., Saleem, A.B., Scho, M.L., Zaharia, A.D., and Carandini, M. (2011). The Detection of Visual Contrast in the Behaving Mouse. *J. Neurosci.* *31*, 11351–11361.
- Carandini, M., and Churchland, A.K. (2013). Probing perceptual decisions in rodents. *Nat Neurosci* *16*, 824–831.
- Charles, L., King, J.-R., and Dehaene, S. (2014). Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *J. Neurosci.* *34*, 1158–1170.
- Churchland, A.K., Kiani, R., Chaudhuri, R., Wang, X.-J., Pouget, A., and Shadlen, M.N. (2011). Variance as a signature of neural computations during decision making. *Neuron* *69*, 818–831.
- Clifford, C.W.G., Arabzadeh, E., and Harris, J.A. (2008). Getting technical about awareness. *Trends Cogn. Sci.* *12*, 54–58.
- Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *5–10*.
- Critchley, H.D., Mathias, C.J., and Dolan, R.J. (2001a). Neural Activity in the Human Brain Relating to Uncertainty and Arousal during Anticipation. *Neuroimage* *29*, 537–545.
- Critchley, H.D., Mathias, C.J., and Dolan, R.J. (2001b). Neural activity in the human brain relating to uncertainty and arousal during anticipation. *Neuron* *29*, 537–545.
- Cromwell, H.C., and Schultz, W. (2003). Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J. Neurophysiol.* *89*, 2823–2838.
- Cul, A. Del, Dehaene, S., Reyes, P., Bravo, E., and Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness.
- Del Cul, a, Dehaene, S., Reyes, P., Bravo, E., and Slachevsky, a (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* *132*, 2531–2540.
- Cury, K.M., and Uchida, N. (2010). Robust odor coding via inhalation-coupled transient activity in the mammalian olfactory bulb. *Neuron* *68*, 570–585.
- Daniel, R., and Pollmann, S. (2012). NeuroImage Striatal activations signal prediction errors on confidence in the absence of external feedback. *Neuroimage* *59*, 3457–3467.
- David, A.S., Bedford, N., Wiffen, B., and Gilleen, J. (2012). Failures of metacognition and lack of insight in neuropsychiatric disorders. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* *367*, 1379–1390.
- Daw, N.D., and Doya, K. (2006). The computational neurobiology of learning and reward. *Curr. Opin. Neurobiol.* *16*, 199–204.
- Dayan, P., and Abbott, L.F. (2005). Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems (MIT Press).

- Dayan, P., Kakade, S., and Montague, P.R. (2000). Learning and selective attention. *Nat. Neurosci.* *3 Suppl*, 1218–1223.
- Deaner, R.O., Khera, A. V, and Platt, M.L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Curr. Biol.* *15*, 543–548.
- Deco, G., Rolls, E.T., Albantakis, L., and Romo, R. (2013). Brain mechanisms for perceptual and reward-related decision-making. *Prog. Neurobiol.* *103*, 194–213.
- Dehaene, S., and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* *70*, 200–227.
- Dehaene, S., Changeux, J., Naccache, L., and Sergent, C. (2006). Conscious , preconscious , and subliminal processing : a testable taxonomy.
- Diamond, M.E., and Arabzadeh, E. (2013). Progress in Neurobiology Whisker sensory system – From receptor to decision. *103*, 28–40.
- Ding, L., and Gold, J.I. (2013). The basal ganglia's contributions to perceptual decision making. *Neuron* *79*, 640–649.
- Doherty, J.P.O., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., and O'Doherty, J.P. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* *38*, 329–337.
- Dorris, M.C., and Glimcher, P.W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* *44*, 365–378.
- Doya, K., Ishii, S., Pouget, A., and Rao, R. (2007). Bayesian Brain - Probabilistic Aproaches to Neural Coding.
- Drugowitsch, J., and Pouget, A. (2012). Probabilistic vs. non-probabilistic approaches to the neurobiology of perceptual decision-making. *Curr. Opin. Neurobiol.* *22*, 963–969.
- Drugowitsch, J., Moreno-Bote, R., Churchland, A.K., Shadlen, M.N., and Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *J. Neurosci.* *32*, 3612–3628.
- Faisal, A.A., Selen, L.P.J., and Wolpert, D.M. (2008). Noise in the nervous system. *Nature* *9*, 292–303.
- Feierstein, C.E., Quirk, M.C., Uchida, N., Sosulski, D.L., and Mainen, Z.F. (2006). Representation of Spatial Goals in Rat Orbitofrontal Cortex. *Neuron* *495*–507.
- Fellows, L.K. (2011). Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. *Ann. N. Y. Acad. Sci.* *1239*, 51–58.
- Felsen, G., and Mainen, Z.F. (2008). Neural Substrates of Sensory-Guided Locomotor Decisions in the Rat Superior Colliculus. *Neuron* *60*, 137–148.
- Felsen, G., and Mainen, Z.F. (2012). Midbrain contributions to sensorimotor decision making. *J. Neurophysiol.* *108*, 135–147.
- Fetsch, C.R., Kiani, R., Newsome, W.T., and Shadlen, M.N. (2014). Effects of Cortical Microstimulation on Confidence in a Perceptual Decision. *Neuron* *83*, 797–804.
- Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* *299*, 1898–1902.
- Fitzgerald, T.H.B., Schwartenbeck, P., and Dolan, R.J. (2014). Reward-Related Activity in Ventral Striatum Is Action Contingent and Modulated by Behavioral Relevance. *34*, 1271–1279.

- Flavell, J. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. *Am. Psychol.* *34*.
- Fleming, S.M., and Dolan, R.J. (2010). Effects of loss aversion on post-decision wagering: implications for measures of awareness. *Conscious. Cogn.* *19*, 352–363.
- Fleming, S.M., and Dolan, R.J. (2012). The neural basis of metacognitive ability. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* *367*, 1338–1349.
- Fleming, S.M., Weil, R.S., Nagy, Z., Dolan, R.J., and Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science* *329*, 1541–1543.
- Frith, C.D. (2012). The role of metacognition in human social interactions. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* *367*, 2213–2223.
- Galvin, S.J., Podd, J. V, Drga, V., and Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon. Bull. Rev.* *10*, 843–876.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychol Rev* *84*, 279–325.
- Gibbon, J., Malapani, C., Dale, C.L., and Gallistel, C. (1997). Toward a neurobiology of temporal cognition: advances and challenges. *Curr. Opin. Neurobiol.* *7*, 170–184.
- Giessel, A.J., and Datta, S.R. (2014). Olfactory maps, circuits and computations. *Curr. Opin. Neurobiol.* *24*, 120–132.
- Glimcher, P.W. (2003). Decisions, uncertainty, and the brain: The science of neuroeconomics (MIT Press).
- Glimcher, P.W., and Rustichini, A. (2004). Neuroeconomics: the consilience of brain and decision. *Science* *306*, 447–452.
- Gold, J.I., and Shadlen, M.N. (2007). The Neural Basis of Decision Making. *Annu. Rev. Neurosci.*
- Green, D., and Swets, J. (1966). Signal Detection Theory and Psychophysics (Wiley).
- Greenspan, R.J. (2007). An Introduction to Nervous System.
- Gremel, C.M., and Costa, R.M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nat. Commun.* *4*, 2264.
- Griffin, D. (1993). The weighing of evidence and the determinants of confidence by Dale Griffin and Amos Tversky. (Stanford Calif.: Stanford Center on Conflict and Negotiation Stanford University).
- Grinband, J., Hirsch, J., and Ferrera, V.P. (2006). A Neural Representation of Categorization Uncertainty in the Human Brain. *Science* (80-.). 757–763.
- Haber, S.N., and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* *35*, 4–26.
- Hampton, R.R. (2001). Rhesus monkeys know when they remember. *Proc. Natl. Acad. Sci. U. S. A.* *98*, 5359–5362.
- Hampton, A.N., Bossaerts, P., and O'Doherty, J.P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* *26*, 8360–8367.

- Hare, T. a, O'Doherty, J., Camerer, C.F., Schultz, W., and 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*, 5623–5630.
- Harvey, N. (1997). Confidence in judgment. *Trends Cogn. Sci.* *1*, 78–82.
- Hebart, M.N., Schriever, Y., Donner, T.H., and Haynes, J.-D. (2014a). The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. *Cereb. Cortex*.
- Herrnstein, R. (1961). Relative and Absolute Strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* *267*–272.
- Higham, P.A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *J. Exp. Psychol. Gen.* *136*, 1–22.
- Horwitz, G.D., and Newsome, W.T. (1999). Separate signals for target selection and movement specification in the superior colliculus. *Science* *284*, 1158–1161.
- Horwitz, G.D., and Newsome, W.T. (2001). Target selection for saccadic eye movements: prelude activity in the superior colliculus during a direction-discrimination task. *J. Neurophysiol.* *86*, 2543–2558.
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., and Camerer, C.F. (2005). Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making. *Science* (80-.). *1680*.
- Ikemoto, S. (2007). Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res. Rev.* *56*, 27–78.
- Insabato, A., Pannunzi, M., Rolls, E.T., and Deco, G. (2010a). Confidence-related decision making. *J. Neurophysiol.* *104*, 539–547.
- Insabato, A., Pannunzi, M., and Deco, G. (2010b). Confidence-related decision-making. *Neuroscience* *104*, 539–547.
- Ito, M., and Doya, K. (2009). Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J. Neurosci.* *29*, 9861–9874.
- Ito, M., and Doya, K. (2011). Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. *Curr. Opin. Neurobiol.* *21*, 368–373.
- Janssen, P., and Shadlen, M.N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nat. Neurosci.* *8*, 234–241.
- Johnson, D.D.P., and Fowler, J.H. (2011). The evolution of overconfidence. *Nature* *477*, 317–320.
- Jones, J.L., Esber, G.R., McDannald, M.A., Gruber, A.J., Hernandez, A., Mirenzi, A., and Schoenbaum, G. (2012). Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science* *338*, 953–956.
- Kennerley, S.W., Behrens, T.E.J., and Wallis, J.D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat. Neurosci.* *14*, 1581–1589.
- Kepecs, A. (2013). The uncertainty of it all. *Nat. Neurosci.* *16*, 660–662.
- Kepecs, A., and Mainen, Z.F. (2012). A computational framework for the study of confidence in humans and animals. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* *367*, 1322–1337.
- Kepecs, A., Uchida, N., Zariwala, H.A., and Mainen, Z.F. (2008). Neural correlates , computation and behavioural impact of decision confidence. *Methods* *455*.

- Kiani, R., and Shadlen, M.N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* *324*, 759–764.
- Kiani, R., Hanks, T.D., and Shadlen, M.N. (2008). Bounded Integration in Parietal Cortex Underlies Decisions Even When Viewing Duration Is Dictated by the Environment. *Analysis* *28*, 3017–3029.
- Knill, D.C., and Pouget, A. (2004). The Bayesian brain : the role of uncertainty in neural coding and computation. *Trends Neurosci.* *27*.
- Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., and Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat. Neurosci.* *16*, 749–755.
- Kording, K. (2007). Decision theory: what “should” the nervous system do? *Science* *318*, 606–610.
- Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J.Z., and Kepcs, a (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* *498*, 363–366.
- Lak, A., Costa, G.M., Romberg, E., Koulakov, A., and Mainen, Z.F. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* *in press*.
- Lau, B., and Glimcher, P.W. (2008). Value representations in the primate striatum during matching behavior. *Neuron* *58*, 451–463.
- Lau, H.C., and Passingham, R.E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 18763–18768.
- Laurent, G., and Fregnac, Y. (2014). Where is the brain in the Human Brain Project? *Nature* *513*, 27–28.
- Lauwereyns, J., Watanabe, K., Coe, B., and Hikosaka, O. (2002). A neural correlate of response bias in monkey caudate nucleus. *Nature* *418*, 413–417.
- Leon, M.I., and Shadlen, M.N. (2003). Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* *38*, 317–327.
- Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci.* *9*, 1432–1438.
- Mainen, Z.F., and Kepcs, A. (2009). Neural representation of behavioral outcomes in the orbitofrontal cortex. *Curr. Opin. Neurobiol.* *19*, 84–91.
- Mainen, Z.F., and Pouget, A. (2014). Put brain project back on course. *Nature* *511*, 2014.
- Mannella, F., Gurney, K., and 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.
- Mar, A.C., Walker, A.L.J., Theobald, D.E., Eagle, D.M., and Robbins, T.W. (2011). Dissociable effects of lesions to orbitofrontal cortex subregions on impulsive choice in the rat. *J. Neurosci.* *31*, 6398–6404.
- Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual Information (San Francisco).
- Martin, J.H., and Ghez, C. (1999). Pharmacological inactivation in the analysis of the central control of movement. *J. Neurosci. Methods* *86*, 145–159.

- Martinez-Garcia, Marina; Insabato, Andrea; Pannunzi, Mario; Pardo-Vazquez, Jose; Acuna, C. and, and Deco, G. (2014). Neural correlates of decision confidence in monkey premotor cortex. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 20167–20172.
- De Martino, B., Fleming, S.M., Garrett, N., and Dolan, R.J. (2013). Confidence in value-based choice. *Nat. Neurosci.* *16*, 105–110.
- Mazurek, M.E. (2003). A Role for Neural Integrators in Perceptual Decision Making. *Cereb. Cortex* *13*, 1257–1269.
- McDannald, M. a, Lucantonio, F., Burke, K. a, Niv, Y., and Schoenbaum, G. (2011). Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J. Neurosci.* *31*, 2700–2705.
- Meer, M.A.A. Van Der, Redish, A.D., and Yin, H.H. (2009). Covert expectation-of-reward in rat ventral striatum at decision points. *3*, 1–15.
- Van der Meer, M. a a, and Redish, a D. (2011). Ventral striatum: a critical look at models of learning and evaluation. *Curr. Opin. Neurobiol.* *21*, 387–392.
- Metcalfe, J. (2008). Evolution of metacognition. In: *Handbook of metamemory and memory*.
- Middlebrooks, P.G., and Sommer, M. a (2012). Neuronal correlates of metacognition in primate frontal cortex. *Neuron* *75*, 517–530.
- Middlebrooks, P.G., and Sommer, M.A. (2011). Metacognition in monkeys during an oculomotor task. *J. Exp. Psychol. Learn. Mem. Cogn.* *37*, 325–337.
- Miura, K., Mainen, Z.F., and Uchida, N. (2012a). Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. *Neuron* *74*, 1087–1098.
- Miura, K., Mainen, Z.F., and Uchida, N. (2012b). Odor representations in olfactory cortex: distributed rate coding and decorrelated population activity. *Neuron* *74*, 1087–1098.
- Moreno-Bote, R. (2010). Decision confidence and uncertainty in diffusion models with partially correlated neuronal integrators. *Neural Comput.* *22*, 1786–1811.
- Morrison, S.E., and Salzman, C.D. (2011). Representations of appetitive and aversive information in the primate orbitofrontal cortex. *Ann. N. Y. Acad. Sci.* *1239*, 59–70.
- Morrison, S.E., Saez, A., Lau, B., and Salzman, C.D. (2011). Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron* *71*, 1127–1140.
- Murakami, M., Vicente, M.I., Costa, G.M., and Mainen, Z.F. (2014). Neural antecedents of self-initiated actions in secondary motor cortex Masayoshi. *Nat. Neurosci.*
- Murthy, V.N. (2011). Olfactory maps in the brain. *Annu. Rev. Neurosci.* *34*, 233–258.
- Navalpakkam, V., Koch, C., Rangel, A., and Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 5232–5237.
- Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. *Nature* *341*, 52–54.
- Nomoto, K., Schultz, W., Watanabe, T., and Sakagami, M. (2010). Temporally Extended Dopamine Responses to Perceptually Demanding Reward-Predictive Stimuli. *30*, 10692–10702.
- Noonan, M.P., Walton, M.E., Behrens, T.E.J., Sallet, J., Buckley, M.J., and Rushworth, M.F.S. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 20547–20552.

- O'Neill, M., and Schultz, W. (2013). Risk prediction error coding in orbitofrontal neurons. *J. Neurosci.* *33*, 15810–15814.
- O'Neill, M., Schultz, W., and Neill, M.O. (2010). Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron* *68*, 789–800.
- Otazu, G.H., Tai, L.-H., Yang, Y., and Zador, A.M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nat. Neurosci.* *12*, 646–654.
- Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model. *Annu. Rev. Neurosci.* *34*, 333–359.
- Padoa-Schioppa, C., and Assad, J.A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.* *11*, 95–102.
- Padoa-Schioppa, C., and Cai, X. (2011). The orbitofrontal cortex and the computation of subjective value: consolidated concepts and new perspectives. *Ann. N. Y. Acad. Sci.* *1239*, 130–137.
- Palmer, J., Huk, A.C., and Shadlen, M.N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.* *5*, 376–404.
- Paton, J.J., Belova, M.A., Morrison, S.E., and Salzman, C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* *439*, 865–870.
- Persaud, N., Mcleod, P., and Cowey, A. (2007). Post-decision wagering objectively measures awareness. *10*, 257–261.
- Platt, M.L., and Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex. *Nature* *400*, 233–238.
- Preuschoff, K., Bossaerts, P., and Quartz, S.R. (2006). Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures. *381–390*.
- Ranade, S.P., and Mainen, Z.F. (2009). Transient firing of dorsal raphe neurons encodes diverse and specific sensory, motor, and reward events. *J. Neurophysiol.* *102*, 3026–3037.
- Rao, R.P.N. (2010). Decision making under uncertainty: a neural model based on partially observable markov decision processes. *Front. Comput. Neurosci.* *4*, 146.
- Raposo, D., Sheppard, J.P., Schrater, P.R., and Churchland, A.K. (2012). Multisensory decision-making in rats and humans. *J. Neurosci.* *32*, 3726–3735.
- Roesch, M.R., Taylor, A.R., and Schoenbaum, G. (2006). Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. *Neuron* *51*, 509–520.
- Roesch, M.R., Singh, T., Brown, P.L., Mullins, S.E., and 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*, 13365–13376.
- Roitman, J.D., and Roitman, M.F. (2010). Risk-preference differentiates orbitofrontal cortex responses to freely chosen reward outcomes. *Eur. J. Neurosci.* *31*, 1492–1500.
- Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* *22*, 9475–9489.
- Rolls, E.T., and Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making. *Prog. Neurobiol.* *86*, 216–244.

- Rolls, E.T., McCabe, C., and Redoute, J. (2008). Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cereb. Cortex* *18*, 652–663.
- Rolls, E.T., Grabenhorst, F., and Deco, G. (2010a). Choice, difficulty, and confidence in the brain. *Neuroimage* *53*, 694–706.
- Rolls, E.T., Grabenhorst, F., and Deco, G. (2010b). Decision-making, errors, and confidence in the brain. *J. Neurophysiol.* *104*, 2359–2374.
- Rounis, E., Maniscalco, B., Rothwell, J.C., Passingham, R.E., and Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci.* *1*, 165–175.
- Rudebeck, P.H., and Murray, E.A. (2008). Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *J. Neurosci.* *28*, 8338–8343.
- Salzman, C.D., and Newsome, W.T. (1994). Neural mechanisms for forming a perceptual decision. *Science* *264*, 231–237.
- Schilman, E. a, Uylings, H.B.M., Galis-de Graaf, Y., Joel, D., and Groenewegen, H.J. (2008). The orbital cortex in rats topographically projects to central parts of the caudate-putamen complex. *Neurosci. Lett.* *432*, 40–45.
- Schoenbaum, G., and Roesch, M. (2005). Orbitofrontal cortex, associative learning, and expectancies. *Neuron* *47*, 633–636.
- Schoenbaum, G., Nugent, S.L., Saddoris, M.P., and Setlow, B. (2002). Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *Neuroreport* *13*, 885–890.
- Schoenbaum, G., Roesch, M.R., Stalnaker, T.A., and Takahashi, Y.K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nat. Rev. Neurosci.* *10*, 885–892.
- Schoenbaum, G., Takahashi, Y., Liu, T.-L., and McDannald, M. a (2011). Does the orbitofrontal cortex signal value? *Ann. N. Y. Acad. Sci.* *1239*, 87–99.
- Schultz, W., Tremblay, L., and Hollerman, J.R. (2000). Reward Processing in Primate Orbitofrontal Cortex and Basal Ganglia. *272–283*.
- Schultz, W., O'Neill, M., Tobler, P.N., and Kobayashi, S. (2011). Neuronal signals for reward risk in frontal cortex. *Ann. N. Y. Acad. Sci.* *1239*, 109–117.
- Schurger, A., and Sher, S. (2008). Awareness, loss aversion, and post-decision wagering. *Trends Cogn. Sci.* *12*, 209–10; author reply 210.
- Shadlen, M.N., and Kiani, R. (2013). Decision making as a window on cognition. *Neuron* *80*, 791–806.
- Shadlen, M.N., and Newsome, W.T. (1996). Motion perception: seeing and deciding. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 628–633.
- Shadlen, M.N., and Newsome, W.T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* *86*, 1916–1936.
- Shidara, M., Aigner, T.G., and Richmond, B.J. (1998). Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J. Neurosci.* *18*, 2613–2625.

- Simmons, J.M., Ravel, S., Shidara, M., and 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*, 376–394.
- Smith, J.D., Schull, J., Strote, J., McGee, K., Egnor, R., and Erb, L. (1995). The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *J. Exp. Psychol. Gen.* *124*, 391–408.
- Smith, J.D., Shields, W.E., Schull, J., and Washburn, D. a (1997). The uncertain response in humans and animals. *Cognition* *62*, 75–97.
- Smith, J.D., Shields, W.E., and Washburn, D. a (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behav. Brain Sci.* *26*, 317–39; discussion 340–73.
- So, N., and Stuphorn, V. (2012). Supplementary eye field encodes reward prediction error. *J. Neurosci.* *32*, 2950–2963.
- Sosulski, D.L., Bloom, M.L., Cutforth, T., Axel, R., and Datta, S.R. (2011). Distinct representations of olfactory information in different cortical centres. *Nature* *472*, 213–216.
- Stettler, D.D., and Axel, R. (2009). Representations of odor in the piriform cortex. *Neuron* *63*, 854–864.
- Sugrue, L.P., Corrado, G.S., and Newsome, W.T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science* *304*, 1782–1787.
- Sugrue, L.P., Corrado, G.S., and Newsome, W.T. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat. Rev. Neurosci.* *6*, 363–375.
- Swanson, L.W. (2003). *Brain Architecture - Understanding the Basic Plan* (Oxford University Press).
- Tai, L.-H., Lee, a M., Benavidez, N., Bonci, A., and Wilbrecht, L. (2012). Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat. Neurosci.* *15*, 1281–1289.
- Takahashi, Y.K., Chang, C.Y., Lucantonio, F., Haney, R.Z., Berg, B. a, Yau, H.-J., Bonci, A., and Schoenbaum, G. (2013). Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron* *80*, 507–518.
- Thomas, G.P. (2002). Research on Sociocultural Influences on Motivation and Learning - The social mediation of metacognition.
- Thompson, J. a, and Felsen, G. (2013). Activity in mouse pedunculopontine tegmental nucleus reflects action and outcome in a decision-making task. *J. Neurophysiol.* *110*, 2817–2829.
- Tobler, P.N., O'Doherty, J.P., Dolan, R.J., and Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.* *97*, 1621–1632.
- Tsujimoto, S., Genovesio, A., and Wise, S.P. (2010). Evaluating self-generated decisions in frontal pole cortex of monkeys. *Nat. Neurosci.* *13*, 120–126.
- Uchida, N., and Mainen, Z.F. (2003). Speed and accuracy of olfactory discrimination in the rat. *Nat. Neurosci.* *6*, 1224–1229.
- Uchida, N., Poo, C., and Haddad, R. (2014). Coding and Transformations in the Olfactory System. *Annu. Rev. Neurosci.*
- Venturini, R., Johnson, H., de Witt, E., and Mainen, Z.F. (2014). Confidence, Waiting time and Feedback. In *79th Cold Spring Harbor Symposium on Quantitative Biology - Cognition*,

- Vickers, D., and Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychol. (Amst).* *50*, 179–197.
- Vilares, I., Howard, J.D., Fernandes, H.L., Gottfried, J. a, and Kording, K.P. (2012). Differential representations of prior and likelihood uncertainty in the human brain. *Curr. Biol.* *22*, 1641–1648.
- Wallis, J.D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* *30*, 31–56.
- Wallis, J.D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat. Neurosci.* *15*, 13–19.
- Walton, M.E., Behrens, T.E.J., Buckley, M.J., Rudebeck, P.H., and Rushworth, M.F.S. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* *65*, 927–939.
- Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. *Neuron* *60*, 215–234.
- Wang, A.Y., Miura, K., and Uchida, N. (2013). The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nat. Neurosci.* *16*, 639–647.
- Wilson, R.I., and Mainen, Z.F. (2006). Early events in olfactory processing. *Annu. Rev. Neurosci.* *29*, 163–201.
- Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., and Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron* *81*, 267–279.
- Yeung, N., and Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* *367*, 1310–1321.
- Yokoyama, O., Miura, N., Watanabe, J., Takemoto, A., Uchida, S., Sugiura, M., Horie, K., Sato, S., Kawashima, R., and Nakamura, K. (2010). Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neurosci. Res.* *68*, 199–206.
- Zariwala, H. a, Kepecs, A., Uchida, N., Hirokawa, J., and Mainen, Z.F. (2013). The limits of deliberation in a perceptual decision task. *Neuron* *78*, 339–351.
- Zemel, R.S., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Comput.* *10*, 403–430.
- Zinck, L., and Lima, S.Q. (2013). Mate choice in *Mus musculus* is relative and dependent on the estrous state. *PLoS One* *8*, e66064.
- Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature* *497*, 482–485.
- Zylberberg, A., Barttfeld, P., and Sigman, M. (2012). The construction of confidence in a perceptual decision. *Front. Integr. Neurosci.* *6*, 79.