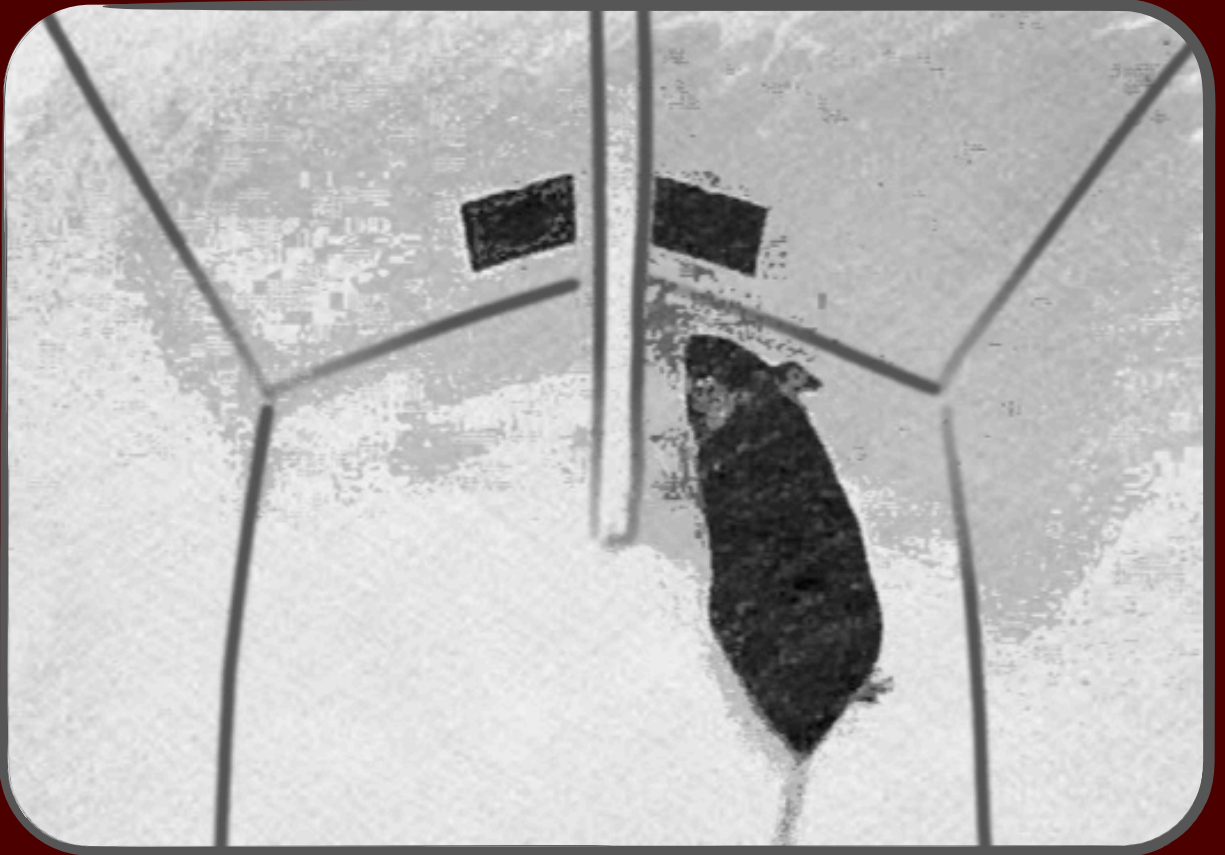


The role of prefrontal cortex and serotonin in foraging decisions

Dario Sarra



Dissertation presented to obtain the Ph.D degree in Neuroscience
Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
December, 2021



UNIVERSIDADE
NOVA
DE LISBOA

Oeiras, December, 2021

The role of prefrontal cortex and serotonin in foraging decisions

Dario Sarra



The role of prefrontal cortex and serotonin in foraging decisions

Dario Sarra

Dissertation presented to obtain the Ph.D degree in Neuroscience

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

Research work coordinated by:



Oeiras, December, 2021



THE ROLE OF PREFRONTAL CORTEX AND SEROTONIN IN FORAGING DECISIONS

Dario Sarra

A dissertation

Presented to the Faculty

Of Universidade Nova de Lisboa

In Candidacy for the Degree

Of Doctor of Philosophy

Supervisor: Zachary F. Mainen

December 2021

Abstract

Deciding whether to give up on a course of action or persist in the hope that our efforts will pay off is a crucial but hard problem to define and optimize. What information is best to take into account? What returns justify the costs? What are the costs and benefits of taking an alternative course of action? Considering all these questions simultaneously seems hard, yet, animals routinely solve this conundrum while foraging.

Using behavioral, computational and neuronal activity manipulation techniques, in this thesis, we investigate how the brain solves this problem. Specifically, we aim to deepen our understanding of two aspects: how the agent represents the environment in order to extract information suitable to guide its foraging decisions, and how serotonin activity can bidirectionally modulate this decision process.

In the first part of the thesis, we focus on finding a heuristic solution that could be adopted by animals to guide foraging behavior under uncertainty. This solution should rely on the same qualitative aspects as the optimal one and, at the same time, be computationally tractable and neurally plausible. Using a newly designed foraging task, we show that, although animals behave in qualitative accordance with the marginal value theorem, a more heuristic proportional hazard model, which includes a leaving probability reset upon reward, better accounts for the behavior.

Next, we focus on the role of serotonin in adjusting the decision to leave. Using a combination of selective pharmacological manipulations and optogenetic manipulations of serotonergic neurons, we show that serotonin can elicit paradoxical behavior effects even

within the same task. Our results support the hypothesis that serotonin's role is to provide the necessary balance between the tendency to exploit or to explore.

In the third part of this work, we adjust our task design to further interrogate the ability of both mice and humans to learn what variables are relevant in the task, and how those variables should be first inferred from the environment, and then used to best guide behavior. Using optogenetic manipulations, we selectively investigate the contribution of different prefrontal cortical areas to the use of inferred variables. Our results provide evidence in support of the hypothesis that mice and humans can infer environment states, and the orbitofrontal cortex plays a crucial role in this ability.

Finally, we investigate the neural cause of the increase in the tendency towards exploitation in foraging that occurs from youth to adulthood in mice. Combining electrophysiology recordings, behavioral assessment and circuit-specific manipulations, we provide evidence to support the hypothesis that the postnatal development of projections from the prefrontal cortex to the dorsal raphe nuclei contributes to the aforementioned behavioral changes.

Taken together, our findings deepen the understanding of how the serotonergic system and prefrontal cortical areas individually contribute to foraging, and how their interactions shape its behavioral and neural development.

Título e Resumo

Decidir entre desistir de um plano de ação e persistir na esperança de que os nossos esforços valham a pena é um problema crucial, mas difícil de definir e otimizar. Qual é a melhor informação a ter em conta? Quais as recompensas que justificam os custos? Quais são os custos e benefícios de adotar um plano de ação alternativo? Considerar todas estas questões em simultâneo parece difícil, no entanto, os animais resolvem rotineiramente este enigma enquanto forrageiam.

Utilizando técnicas comportamentais, computacionais e de manipulação da atividade neuronal, nesta tese investigamos como o cérebro resolve este problema. Especificamente, o nosso objetivo é aprofundar a nossa compreensão de dois aspetos: como o agente representa o ambiente para extrair informação capaz de o guiar nas suas decisões de forrageamento e como a atividade da serotonina pode modular bidirecionalmente o processo de decisão.

Na primeira parte da tese, concentrámo-nos em encontrar uma solução heurística que possa ser adotada pelos animais para guiar o comportamento de forrageamento sob incerteza. Esta solução deve basear-se nos mesmos aspetos qualitativos que a solução ótima e, ao mesmo tempo, ser computacionalmente gerível e neuralmente plausível. Utilizando uma nova tarefa de forrageamento, mostramos que, embora os animais se comportem de acordo com o teorema do valor marginal, um modelo mais heurístico de risco proporcional, que inclui um reset da probabilidade de partir no momento da recompensa, explica melhor o comportamento.

Em seguida, concentramo-nos no papel da serotonina no ajuste da decisão de partir. Utilizando uma combinação de manipulações farmacológicas seletivas e manipulações optogenéticas de neurónios serotoninérgicos, mostramos que a serotonina pode provocar efeitos comportamentais paradoxais mesmo dentro da mesma tarefa. Os nossos resultados apoiam a hipótese de que o papel da serotonina é fornecer o equilíbrio necessário entre a tendência a tirar partido da opção atual ou explorar alternativas.

Na terceira parte deste trabalho, ajustamos a nossa tarefa para examinar ainda mais a capacidade tanto dos ratos como dos humanos de aprender quais variáveis são relevantes na tarefa, e como essas variáveis devem ser primeiro inferidas do ambiente, e depois utilizadas para melhor orientar o comportamento. Através de manipulações optogenéticas, investigamos seletivamente a contribuição de diferentes áreas corticais pré-frontais para a utilização de variáveis inferidas. Os nossos resultados apoiam a hipótese de que ratos e humanos podem inferir estados ambientais, e o córtex orbitofrontal desempenha um papel crucial nesta capacidade.

Finalmente, investigamos a causa neural do aumento da tendência para a exploração no forrageamento que ocorre desde a juventude até à idade adulta em ratos. Combinando registos eletrofisiológicos, avaliação comportamental e manipulações específicas de circuitos, verificamos que o desenvolvimento pós-natal de projeções desde o córtex pré-frontal até aos núcleos dorsais da rafe está subjacente às alterações comportamentais acima mencionadas.

Em conjunto, as nossas descobertas aprofundam a compreensão de como o sistema serotoninérgico e as áreas corticais pré-frontais contribuem individualmente para o forrageamento, e como

as suas interações moldam o seu desenvolvimento comportamental e neural.

Author Contributions

This thesis was written by Dario Sarra. Detailed contributions are provided per chapter in the “Author contributions” section of each chapter.

Chapter 2 is a published manuscript, in collaboration with Eran Lottem, Dhruba Banerjee, Pietro Vertechi, Dario Sarra, Matthijs Oude Lohuis & Zachary F. Mainen.

Chapter 3 is an ongoing work, in collaboration with Dario Sarra, Eran Lottem and Zachary F. Mainen.

Chapter 4 is a published manuscript in collaboration with Pietro Vertechi, Eran Lottem, Dario Sarra, Beatriz Godinho, Isaac Treves, Tiago Quendera, Matthijs Nicolai Oude Lohuis, Zachary F. Mainen.

Chapter 5 is an ongoing work in collaboration with Nicolas Gutierrez-Castellanos, Dario Sarra, Beatriz S. Godinho and Zachary F. Mainen.

Financial Support

This work was carried out under the International Neuroscience and Physiology Doctoral Program (INPDP), funded by the Portuguese FCT (Fundação para a Ciência e a Tecnologia, FCT Bolsa SFRH/BD/132172/2017) and the Fundação Champalimaud.

Table of content

Abstract	3
Título e Resumo	5
Author Contributions	7
Financial Support	7
Table of content	8
Chapter 1: Introduction	1
The foraging dilemma and the optimal foraging theory	2
References	13
Executive functions and the prefrontal cortex	16
References	23
Persistence and flexibility, and their modulation by serotonin	30
References	40
Objectives and organization of the thesis	46
Chapter 2: Activation of serotonin neurons promotes active persistence in a probabilistic foraging task.	47
Introduction	49
Results	50
Mouse behavior in a probabilistic foraging task	50
Activation of DRN neurons prolongs active exploitation	51
Activation of DRN neurons biases leaving but not travel time	51
The proportional hazards model of probabilistic foraging	52
Vigor of nose-poking reports the hazard rate of leaving	53
Discussion	54
Materials and Methods	56
References	57
Author Contributions	59
Chapter 3: Serotonin elicits paradoxical effects on behavioral persistence and flexibility in foraging decisions	60
Introduction	61
Results	64

Animals engaging in a foraging task adopt a heuristic in line with the marginal value theorem solution	64
Foraging task engages the 5-HT receptome in a similar manner to reversal and steady tasks	69
Sustained and transient serotonin produce paradoxical effects on foraging decisions	73
Discussion	79
Materials and Methods	83
References	87
Author Contributions	93
Supplementaries	93
Chapter 4: Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas	101
Introduction	102
Results	103
A Probabilistic Foraging Task Can Dissociate Value or Evidence Accumulation	103
Mice Accumulate Evidence and Not Rewards	104
Accumulation of Evidence Is Tuned to Task Parameters	106
Humans Perform Inference and Tune Behavior to Task Parameters	107
OFC, but Not ACC, Is Necessary for the Correct Inference Process	107
Discussion	109
References	111
Author contributions	111
Materials and Methods	113
Supplementaries	119
Chapter 5: Maturation of prefrontal input to dorsal raphe increases behavioral persistence in mice	123
Summary	123
Introduction	124
Results	127
Cortical top-down input over the dorsal raphe matures in the transition between adolescence and adulthood in mice.	128

Baseline persistence correlates with the maturation of cortico-raphe input in the transition between adolescence and adulthood in mice.	132
mPFC-DRN pathway ablation in adult mice recapitulates juvenile behavioral features.	135
Discussion	140
Materials and Methods	145
References	152
Author contributions	163
Supplementaries	164
General discussion	172
Main findings	172
A hypothetical circuitry	173
Testing circuitry hypothesis	176
Final remarks	177
References	178

When life gives you lemon,

make lemonade.

Elbert Hubbard

Chapter 1: Introduction

Our decisions define who we are in the eyes of others, but for the people around us, the process that led up to them is an inaccessible realm hidden behind our actions. Despite their critical importance, the thoughts, considerations, and doubts that precede a choice go utterly undetected by others. So much so that a large part of personal conflicts can be resolved, or at least eased, just by sharing and making explicit the decision process behind each other's actions. Engaging in decisions is an omnipresent phenomenon that we often forget to reflect on how such a process occurs, much like breathing. Yet, most fields of human knowledge have eventually grappled with questions about how choices are made. Sociologists and anthropologists investigate how society and culture influence and derive from local decisions. Economists examine how we ought to make decisions in a market and how those choices shape individual and market behavior. Modern-day Artificial Intelligence research tries to develop agents increasingly capable of making independent decisions and perhaps, shortly, capable of generalizing this ability. In Psychiatry and Psychology, detailed descriptions of neurotypical decision-making are constantly expanded and at the core of many diagnostic and therapeutic tools. Neuroscience aims to understand the neural mechanisms underlying simple and complex decisions. Its goal is to map function to brain circuits and understand the general principles of how circuits are organized, what computations they perform, and how they ultimately give rise to decisions.

This thesis takes a systems neuroscience approach. To study the neural and computational mechanisms underlying foraging decisions, we combine computational descriptions of behavior, selective neural manipulations, and ethologically relevant behavioral tasks. This chapter

will review the relevant literature that motivated and grounded the later presented work, both conceptually and technically. First, we introduce the dilemma at the core of foraging decision-making and outline the developments that led to our present-day normative theories of foraging. Next, we review the role of different prefrontal cortical regions in executive functions and their involvement in foraging. We then discuss persistence and flexibility, two core tendencies that need to be balanced to achieve optimal foraging, and how the neuromodulator serotonin modulates them. Finally, I present the objectives of the thesis and outline its structure.

The foraging dilemma and the optimal foraging theory

All living beings need to procure nutrients to sustain their organisms. Some have the luxury to extract sustenance directly from their surrounding environment, like plants and corals, but others, like us, need to move, find nourishment and collect it. When feeding meets motion, we have foraging, “the action of going from place to place searching for things that you can eat” (Cambridge dictionary). At first glance, it may appear a simple problem, but zooming in, we can start seeing its complexity. Given the dynamic and uncertain nature of natural environments, how does an organism decide when to spend energy to exploit a current option or leave to explore the environment for a better one? This dilemma is at the core of foraging decision-making: do I stay? Or do I switch?

Foraging is a problem that evolution has faced in countless variations. Given its crucial importance, it is perhaps not surprising that multiple science domains have been interested in foraging. What began

as a concern regarding solely ecology has become a general framework. In part, this expansion is explained by the fact that when foraging is studied at the individual level, it is relevant to many disciplines, including biology, anthropology, economics, and psychology and sub-branches such as ethology, animal learning, and sociobiology.

In reviewing the literature on foraging, multiple roads could be taken. Since technical reviews are abundant (Kamil, Krebs, and Pulliam, 1987; Hayden and Walton, 2014; Calhoun and Hayden, 2015) and certainly far more complete than what we could expect to achieve in this introductory chapter, we chose to take a historical perspective focusing on the development of Optimal Foraging Theory (OFT). This choice was made to serve the reader and us with an initial high-level perspective and highlight the general dynamics that characterize the interaction of empirical and theoretical approaches.

OFT aims to explain/describe algorithmically which factors determine foraging choices and is at the core of the computational approach that we took to investigate the neural basis of decision making. At the core of this approach is the critical idea that animals can be considered economical decision-makers, that evolution has selected to maximize benefits and minimize costs (Hayden and Walton, 2014). The concept of Optimal foraging was nearly absent before it exploded around the late 1960s and 1970s, with the foundation of Optimal Foraging Ecology.

The first step was the introduction of the “net energy maximization per unit time.” This terminology to describe foraging appeared in 1966 with the back-to-back publication in *American Naturalist* of two papers, one by MacArthur and Pianka, the other by Emlen (MacArthur and Pianka, 1966; Emlen, 1966). In this early phase,

optimal foraging theory was solely focused on explaining diets and food preferences *per se*. However, the idea that one can interpret evolutionary adaptation, such as foraging, as an economic agent tasked to optimize benefit over costs was a far-reaching, revolutionary approach.

The critical intuition was that fitness behaves mathematically similarly to an economic decision-maker, and, therefore, we can study behavior as a mathematical optimization problem (McNamara and Houston, 1986; Stephens and Krebs, 1986, Hayden and Walton 2014). MacArthur and Pianka's paper is a graphical formulation of the general qualitative properties of foraging diet decisions, later called "the fundamental theorem of OFT" (Charnov, cited in Schoener, 1987). The authors explicitly declare that their approach looks at how foraging decisions are made "if species acted in the most economical fashion" (MacArthur and Pianka, 1966), arguing that evolutionary pressure would have "hopefully achieved such optimal allocation of time and energy expenditure" (MacArthur and Pianka, 1966). Despite presenting all the qualitative properties of foraging, the authors did not seem to have realized at that point the vast extent of application their model could offer. In particular, they did not mention any aspect of individual behavior. The applications proposed focused on population and evolutionary ecology, such as the effects of the arrival of an invading species. This characteristic is not a peculiarity of their approach but a feature of their time ecology field. General ecology texts pre-1966 rarely contained any reference to feeding behavior and no specialized section on individual ecology (Schoener, 1987).

In contrast to MacArthur and Pianka, the prospective applications of Emlen's paper were much closer to behavioral ecology. In introducing his work, Emlen refers to contemporary research that had investigated

the phenomenon of food preference from a behavioral perspective (i.e., Barnett (1953) on the importance of learning in rats' diet), but also highlights the lack of any attempt to quantitatively relate diet preference to the concept of energy/calories gained per time unit. Despite this, the mathematical formulation proposed was not as effective as the ones that followed. Specifically, Emlen's probabilistic formulation was untenable since it hiddenly implied that the forager is prescient of the prey encounters (Schoener, 1971). Nonetheless, the paper and its discussion offered stimulating biological consideration to the field that substantially shifted the framing of optimal foraging towards what we are accustomed to today (Schoener, 1987). Ultimately, these two papers jointly spelled out a new field of behavioral investigation together with the optimization problem that underlies it: a foraging activity should be increased as long as the "resulting gain in time spent per unit food exceeds the loss" (MacArthur and Pianka, 1966). Schematically, both works established three key principles of OFT: i) foraging can be seen as an optimization issue, solvable adopting a single decision variable that integrates cost and benefit of foraging options, ii) choices can be predicted comparing a foreground option against a background option, and iii) foraging is a dilemma that animals face and solve in their environment, and as such, ecological and ethological considerations are central to it (Hayden and Walton 2014). The core issue to solve optimal foraging was to define the contingency model that describes what constitutes an energy and time budget and how actions increase or decrease these components.

The first algebraic solution of the contingency model appeared in Schoener (1969a). Interestingly, even in this case, the author's intention was not a quantitative behavior description. The formulation was a by-product of modeling optimal body size for predators that employ an active search strategy compared to sit-and-wait predators. The

contingency equation was not explicitly written down in the article, but the time to minimize a metabolic need was minimized for other purposes. For the author, this was just a submodel of a study concerned with understanding the evolution of size differences between species. Consequently, once again, the properties of the contingency model and the optimal foraging aspect were not considered. The formal equation only appeared in the later work by Schoener (1971) as the following (notation as in Schoener 1974b):

$$Y_T = \frac{\sum_{diet} p_i e_i - C_s T_s}{\sum_{diet} p_i t_i + T_s}$$

Where:

- Y_T is the net energy gained per unit of time
- e_i is the net energy for a single foraging item of Type i (potential energy minus exploitation costs)
- t_i is the exploitation costs for a single item of Type i
- p_i is the frequency of Type i in the environment
- T_s is the mean search time per item of available/encountered food
- C_s is the search energy cost per unit search time

Once more, the major mathematical properties of the contingency model were not made explicit. Nevertheless, this time the framing of the paper directly addressed the behavioral relevance of foraging, as it is well exemplified by the title “Theory of feeding strategies.”

Between 1973 and 1974, several quantitative solutions to the problem were proposed by multiple authors, each time with a different

notation and sometimes in ways that suggest partial or total ignorance of previous formulations (Charnov and Orian, 1973; Timin, 1973; Pearson, 1974; Maynard Smith, 1974; Pulliam, 1974; Werner and Hall, 1974; Schoener, 1974 which reported more explicitly and completely the equation in Schoener, 1971). Charnov's body of work excelled in clarity and simplicity, culminating the purely theoretical trajectory of that time with the 1976 American Naturalist paper, titled "Optimal foraging: Attack strategy of a Mantid" (Charnov, 1976a). What sets aside Charnov's work from the other formulation is the presence of the first clear enunciation of the mathematical properties of the contingency model.

- To compute the optimal diet, food type options are ranked in decreasing order according to their energy per unit time value (e/t). Beginning with only the largest e/t food type, the diet value is computed recursively, adding the next ranked item at each step. The process continues until the e/t of the food type x is smaller than the e/t of the diet containing only the higher rank items.
- The larger the food availability, the more specialized (less food types) the optimal diet.
- A food type inclusion is independent of its abundance in the environment. Indeed it solely depends on the absolute abundance of food types with larger e/t values.
- There is no partial preference for food type. One kind of food is either always or never taken when encountered.

The explicit enunciation of these properties propelled the diffusion of Optimal Foraging Theory (Fig. 1), offering the possibility to design animal experiments aiming to directly test the model predictions, such as the Mantid test presented in the same article (Charnov, 1976) and on many other species (e.g., Pyke, Pulliam and Charnov, 1977;

Krebs, Kacelnik and Taylor, 1978). Furthermore, having access to such predictions allowed other fields to test the explanatory power of the optimal diet choice framework, in particular, and OFT more in general.

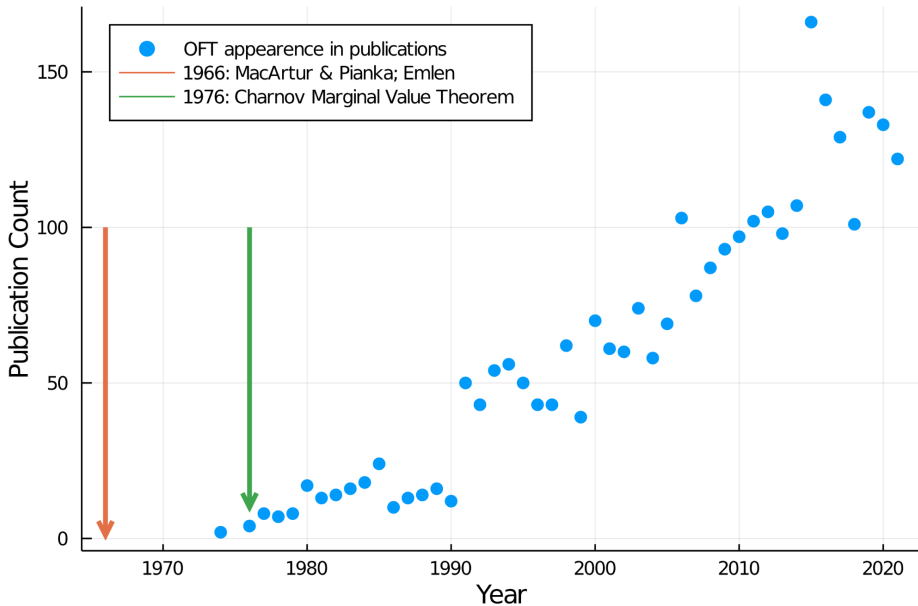


Figure 1. Growth of the optimal foraging theory research. Scatter plot showing the amount of publication containing either “optimal foraging” or “optimal feeding” in a search on Web of Science (Clarivate). Depending on the original manuscript availability, the search may or may not include the full text, as evident from the lack of data from 1966, when MacArthur & Pianka, and Emlen papers came out.

In “A Brief History of Foraging Ecology” (1987), Shoener reflects on the possible causes that led so many scientists, who were interested in the same issue, to derive the same solution independently. He speculates that there could be four possible reasons. We find it interesting to revisit them in this thesis to consider if the work presented here could have suffered from similar issues. First, although not connected, most derivations had the same starting point, the MacArthur and Pianka’s (1966) graphical model. Second, the use of different notations reflected the different considerations the authors wanted to

emphasize, yet it led to erroneously distinguishing the results between each other. In a personal exchange, Pulliam admits to Schoener that his 1974's model essentially restated Schoener's 1971 formulation (cited in Shoener, 1987). In private communication with Schoener, Charnov "referred to the entire phenomenon as rediscovering the wheel" (as reported in Schoener, 1987). Third, Shoener suggests that a part of the proposed equations were written in a not very useful form, that is to say, not highlighting the generality of the case. Finally, the author argues that contemporary ecology theoreticians partially ignored each other's papers, slowing down progress in the field.

Although optimal diet theory was firmly established around 1974, this did not emerge from solely theoretical inceptions. Indeed, several pre-1974 studies provided some empirical evidence of what are later on going to be the predictions of the theoretical formulation. Shoener (1987) provided two examples of optimal diet theorem properties that were previously proved empirically: i) In line with property 2, Ivlev (1961) showed that the diet of fishes become more specialized as the food abundance increase (as reported by Bardach, Ivlev, and Scott translation, 1962), ii) as per property 3, Orians (1966) showed that in time of the day when the best prey option is available yellow-headed blackbirds (*Xanthocephalus xanthocephalus*) almost exclusively catch those, while less valuable prey are caught only in periods when the most valuable options are absent. More examples of a posteriori evidence for the optimal diet theory in the pre-1974 ecology literature are easily found. Young (1946) performed extensive food preference training in rats, concluding that training and habit only have an effect if options presented have comparable palatability, but none if one option has superior palatability. Furthermore, he showed that the animals' metabolic needs determine food palatability. Barnett (1956) documented the

behavioral repertoire of foraging rats and, as a sort of side observation, reported that rat diet is highly specialized when multiple options are available and abundant. Similar results were reported by Thompson's (1965) extensive plant preference experiment on Meadow Vole (*Microtus pennsylvanicus*). Moreover, Thompson ranked the voles' diet according to the voles' preferred items and highlighted the correlation between item's preference and caloric value, as predicted per optimal diet. In summary, the theoretical framework effectively emerged from the gathering of numerous independent empirical studies reporting both parts of the assumptions (maximization of gained energy) and part of the predictions (specialization of diet correlates with abundance) of the optimal diet model.

What did the theoretical framework add to foraging ecology then? First and foremost, it linked together in a cause and effect relationship observations that were otherwise loosely related. Our ability to recognize empirical evidence as part of the foraging strategy is effectively a hindsight product of having an optimal foraging theory. Second, while before the establishment of OFT, the causes of the empirical observation were confined to discussions, thanks to the theoretical framework's predictions, the empirical studies could then directly test which factors played a role in shaping diet. Indeed, in 1974, Werner and Hall gave the first example of direct experimental testing postulated from the optimal diet model. Testing bluegill sunfish's diet in three conditions of preys' abundance, they found that animals behaved in line with the predictions not only at a qualitative level but quantitative too (Werner and Hall, 1974). Furthermore, OFT propelled the somewhat limited empirical research on foraging towards an across-the-board expansion in interest and participation. This new focus on individual behavioral choices during foraging eventually led to the establishment of behavioral ecology as an

independent discipline. To appreciate the relevance of this impulse, one has to consider that, before the OFT establishment, ecology texts were largely devoid of any section dedicated to individual ecology (physiological, evolutionary, and behavioral, Shoener, 1987). Finally, having access to the equation solving the optimal diet issue enabled the solution of other foraging questions thanks to the simplicity of extrapolating formal analogy (e.g., Kacelnik, 1984).

A direct example of this extension is the formulation of the optimal patch choice theory. This theoretical framework addresses the implications of dealing with foraging patch options rather than food types. In a nutshell, the solution can be found by applying the optimal diet equation considering the energy over time (e/t) gained from different foraging patches rather than different food types. From here, a distinction can be derived between competition over food type or patches' richness, for example, in the case of a new species invading an environment. On the one hand, when competition occurs over food types, the overall abundance will decrease, leading to less specialized diets and, therefore, more overlap between residents and invading species' diets. On the other hand, when the competition occurs over patches, the e/t of specific patches would decrease, potentially leading to abandoning such options from the diet and reducing the overlap. This set of predictions is called "The Compression Hypothesis" and was already an objective of MacArthur and Pianka's 1966 paper and received early confirmation, for instance, in Werner and Hall's work on habitat shift (1976, 1977, 1979). Another prominent offshoot of the optimal foraging diet, particularly appreciated in the behavioral neuroscience field, deals with the issue of optimal departure time from patches, which led to the formulation of the Marginal Value Theorem (MVT; Charnov, 1976b). The fundamental solution of Charnov's theory is to replace the fixed energy per patch, as

previously formulated, with a time-dependent version of it. From this, it is possible to derive a marginal capture rate and find that the optimal leaving time corresponds to when the patch gain rate is equal to the average environment's gain rate (a more detailed discussion is presented in Chapter 2).

In conclusion, the development of the OFT is a key example of the benefit of combining experimental and theoretical approaches to explain behavior. In this case (if not in most cases), the theoretical framing's invaluable contribution is the conversion of case-specific observation into an explanatory mechanism, from which testable predictions can be extracted. Furthermore, it shows how a theoretical framework that originated to address a specific issue can be expanded to answer questions formally analogous to the original one.

Curiously, the studies in mice presented in this thesis are related in a manner similar to the OFT experimental and theoretical historical interactions. For instance, in Chapter 2, we employed a foraging task with depleting reward probability. Our behavioral assessment revealed that mice appeared to adjust the number of attempts performed in a foraging patch in a manner that suggested knowledge of this hidden decaying probability. Inspired by these empirical results, in Chapter 4, we redesign our foraging task to test a priori whether animals could learn to represent the environment states using such hidden variables. In turn, thanks to the developed theoretical framework, we were able to experimentally determine the distinctive contribution of the Anterior Cingulate Cortex and Orbitofrontal Cortex in our foraging behavior.

Another instance where we benefit from using a previously developed theoretical framework is the research performed in Chapter 5. To investigate the behavioral impact of the postnatal development of the

prefrontal cortex (PFC) to dorsal raphe nucleus (DRN) projections, we were able to exploit our previous theoretical understanding of foraging and capture behavioral phenotype changes linked to this phenomenon.

Overall our goal in giving this historical review is to propose that the curiosity and dynamical interactions that propelled the OFT field to its contemporary relevance are the same ones that shaped the development of the thesis work presented here.

References

- Bardach, J. E., Ivlev, V. S., & Scott, D. (1962). Experimental Ecology of the Feeding of Fishes [Article]. *Copeia*, 1962(1), 234. <https://doi.org/10.2307/1439521>
- Barnett, S. A., & Spencer, M. M. (1953). Experiments on the food preferences of wild rats (*Rattus norvegicus* Berkenhout). *Journal of Hygiene*, 51(1), 16–34. <https://doi.org/10.1017/S002217240001545X>
- Barnett, S. A. (1956). Behaviour components in the feeding of wild and laboratory rats. *Behaviour*, 176(9), 24–42.
- Calhoun, A. J., & Hayden, B. Y. (2015). The foraging brain. *Current Opinion in Behavioral Sciences*, 5, 24–31. <https://doi.org/10.1016/j.cobeha.2015.07.003>
- Charnov, E. L., & Orians, G. H. (1973). Optimal foraging: some theoretical explorations (24th–2nd–2006th ed.). UNM Digital Repository, University of Mexico. https://digitalrepository.unm.edu/biol_fsp/45
- Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, 9(2), 129–136. [https://doi.org/10.1016/0040-5809\(76\)90040-X](https://doi.org/10.1016/0040-5809(76)90040-X)
- Charnov, E. L. (1976). Optimal Foraging: Attack Strategy of a Mantid. *The American Naturalist*, 110(971), 141–151. <https://doi.org/10.1086/283054>
- Emlen, J. M. (1966). The Role of Time and Energy in Food Preference. *The American Naturalist*, 100(916), 611–617. <https://doi.org/10.1086/282455>
- Hayden, B. Y., & Walton, M. E. (2014). Neuroscience of foraging. *Frontiers in Neuroscience*, 8(3), 358–378. <https://doi.org/10.3389/fnins.2014.00081>
- Kacelnik, A. (1984). Central Place Foraging in Starlings (*Sturnus vulgaris*). I. Patch Residence Time. *The Journal of Animal Ecology*, 53(1), 283. <https://doi.org/10.2307/4357>

- Kamil, A. C., Krebs, J. R. K., & Pulliam, H. R. (1987). *Foraging Behavior* (A. C. Kamil, J. R. Krebs, & H. R. Pulliam (eds.)). Springer US.
<https://doi.org/10.1007/978-1-4613-1839-2>
- Krebs, J. R., Kacelnik, A., & Taylor, P. (1978). Test of optimal sampling by foraging great tits. *Nature*, 275(5675), 27–31. <https://doi.org/10.1038/275027a0>
- MacArthur, R. H., & Pianka, E. R. (1966). On Optimal Use of a Patchy Environment. *The American Naturalist*, 100(916), 603–609.
<https://doi.org/10.1086/282454>
- McNamara, J. M., & Houston, A. I. (1986). The Common Currency for Behavioral Decisions. *The American Naturalist*, 127(3), 358–378.
<https://doi.org/10.1086/284489>
- Orians, G. H. (1966). Food of Nestling Yellow-Headed Blackbirds, Cariboo Parklands, British Columbia. *The Condor*, 68(4), 321–337.
<https://doi.org/10.2307/1365448>
- Pulliam, H. R. (1974). On the Theory of Optimal Diets. *The American Naturalist*, 108(959), 59–74. <https://doi.org/10.1086/282885>
- Pyke, G. H., Pulliam, H. R., & Charnov, E. L. (1977). OPTIMAL FORAGING : A SELECTIVE REVIEW OF THEORY AND TESTS. *The Quarterly Review of Biology*, 52(2), 137–154. <https://doi.org/10.1086/409852>
- Schoener, T. W. (1987). A Brief History of Optimal Foraging Ecology. In A. C. Kamil, J. R. Krebs, & H. R. Pulliam (Eds.), *Foraging Behavior* (Vol. 16, Issues 1–2, pp. 5–67). Springer US. <https://doi.org/10.1007/978-1-4613-1839-2>
- Schoener, T. W. (1969). Models of Optimal Size for Solitary Predators. *The American Naturalist*, 103(931), 277–313. <https://www.jstor.org/stable/2459329>
- Schoener, T. W. (1971). Theory of Feeding Strategies. *Annual Review of Ecology and Systematics*, 2(1), 369–404.
<https://doi.org/10.1146/annurev.es.02.110171.002101>
- Schoener, T. W. (1974). The Compression Hypothesis and Temporal Resource Partitioning. *Proceedings of the National Academy of Sciences*, 71(10), 4169–4172. <https://doi.org/10.1073/pnas.71.10.416>
- Smith, J. M. (1974). *MODELS IN ECOLOGY*.
[https://batrachos.com/sites/default/files/pictures/Books/Maynard Smith_1976_Models in Ecology.pdf](https://batrachos.com/sites/default/files/pictures/Books/Maynard_Smith_1976_Models%20in%20Ecology.pdf)
- Stephens, D. W., & Krebs, J. R. (1987). *Foraging Behavior*. In A. C. Kamil, J. R. Krebs, & H. R. Pulliam (Eds.), *Foraging Theory*. Springer US.
<https://doi.org/10.1007/978-1-4613-1839-2>

- Thompson, D. Q. (1965). Food Preferences of the Meadow Vole (*Microtus pennsylvanicus*) in Relation to Habitat Affinities. *American Midland Naturalist*, 74(1), 76. <https://doi.org/10.2307/2423121>
- Timin, M. E. (1973). A multispecies consumption model. *Mathematical Biosciences*, 16(1–2), 59–66. [https://doi.org/10.1016/0025-5564\(73\)90044-8](https://doi.org/10.1016/0025-5564(73)90044-8)
- Werner, E. E., & Hall, D. J. (1976). Niche Shifts in Sunfishes: Experimental Evidence and Significance. *Science*, 191(4225), 404–406. <https://doi.org/10.1126/science.1246626>
- Werner, E. E., & Hall, D. J. (1974). Optimal Foraging and the Size Selection of Prey by the Bluegill Sunfish (*Lepomis Macrochirus*). *Ecology*, 55(5), 1042–1052. <https://doi.org/10.2307/194035>
- Werner, E. E., & Hall, D. J. (1979). Foraging Efficiency and Habitat Switching in Competing Sunfishes. *Ecology*, 60(2), 256–264. <https://doi.org/10.2307/1937653>
- Werner, E. E., & Hall, D. J. (1977). Competition and Habitat Shift in Two Sunfishes (*Centrarchidae*). *Ecology*, 58(4), 869–876. <https://doi.org/10.2307/1936222>
- Young, P. T. (1945). Studies of food preference, appetite and dietary habit. IV. The balance between hunger and thirst. *Journal of Comparative Psychology*, 38(3), 135–174. <https://doi.org/10.103>

Executive functions and the prefrontal cortex

Since its inception, the marginal value theorem (MVT; Charnov, 1976) has provided a remarkable level of prediction at the qualitative (e.g. Pyke, Pulliam and Charnov, 1977) and, occasionally, at the quantitative level (e.g. Werner and Hall, 1974). Furthermore, the simplicity of its basic algorithm enabled its application to a large set of analogous problems, becoming effectively a standard approach to solve stay-leave decisions in several fields (Parker, 1991; Bird et al., 2009; Walters et al., 2016; Vandeviver, 2021). Part of the power of the MVT formulation comes from combining intake and costs into a single net variable, then comparing the foreground and background net gains. While this is a very effective mathematical description of the behavioral phenomenology, its simplicity disguises the complexity of the computational efforts performed by the nervous system during foraging. The variety of functions necessary to hypothetically execute the MVT algorithms is large and requires coordination between its parts. For instance, just the estimation of the net gain of the foreground option requires combining multiple information. How much time has elapsed since reaching this patch? How nutritious are the items collected? How hard is it to collect one? In addition, while in modelling we solve these questions sequentially, the brain has to address these and more at the same time.

In general, the ability to combine and contextualise information, coordinating the activity of several neural systems (Roberts, 1998; Roberts et al., 1998; Miyake and Shah 1999a,b) is ascribed to the so called executive functions. Various authors have proposed detailed definitions that can broadly be divided in two categories. Those

specifying what constitutes (Burgess, 1997; Baddeley and Della Sala, 1998; Pineda et al., 1998) or what requires executive functions (Shallice and Burgess, 1991; Rabbit, 1997; Perner and Lang, 1999; Smith and Jonides, 1999). Funahashi and Andreau (2013) summarise the common features specified by previous authors in five operations attributed to executive functions:

- control of attention (switching attention from one source to another or focusing attention on one source);
- the temporal organization of behavior;
- the planning or scheduling of complex tasks to accomplish a future goal;
- the capacity to access and manipulate information stored in long-term memory;
- the monitoring of current internal and external states.

In this section, we focus on a few key papers that shaped the thinking behind this thesis's work on prefrontal cortex and foraging. In particular, in how executive functions enable optimal foraging strategy and the evidence supporting the hypothesis that prefrontal cortical (PFC) areas play a major and specific role in action selection and scheduling during foraging. For a broader and more extensive review, see e.g. Rudebeck and Izquierdo (2021) and Roberts et al. (1998). We will proceed by first breaking down the MVT computation into smaller parts and looking at the research addressing where in the brain these computation might occur.

Before we dive into the behavioural literature, a few notes on comparative brain anatomy. Although there is no overarching consensus on how to define PFC across different species, in this thesis, we aligned with the proposal of Uylings and van Eden (1990), which delimits

rodents' prefrontal cortex to those areas comparable to the primate prefrontal cortex in terms of thalamic reciprocal connections, corticocortical connections, and functional aspects (Uylings and van Eden, 1990). Other examples of possible criterias to define rodent areas equivalent to primates PFC include for instance cytoarchitecture aspects, which stress the lack of granular cortex in the rodent PFC (Preuss, 1995; Wise 2008). Furthermore, it should be stressed that independently of the strength of the adopted anatomical criteria, functional differences at the behavioral level are still present between primates and rodents (Wallis et al., 2017), and between human and non-human primates (Roberts et al., 1998). Therefore, caution is needed when making comparisons. For a more recent extensive discussion of rodents and primates comparative brain anatomy see Mars, Jbabdi and Rushworth, 2021.

One important quantity to begin our function-based investigation of the neural basis of foraging is the average reward rate. Indeed, what distinguishes foraging decisions is that the current foreground value is weighted against the environment background option, rather than contrasting the values of multiple options at the same time like in neuroeconomics settings (Hayden and Walton, 2014). To estimate and use this quantity, several elements are necessary. One needs to be able to integrate information over a time window, to maintain such information in the working memory, and to constantly update it to capture changes in the environment. Coordinating such a calculation is likely to involve executive functions. A likely candidate PFC area to engage in the computation of the average reward reward is the anterior cingulate cortex (ACC). In an influential paper published in *Science*, Kolling, Behrens, Mars and Rushworth used functional magnetic resonance imaging (fMRI) to investigate the neural mechanisms mediating foraging

in humans (Kolling et al., 2012). Subjects navigated a virtual environment and were tasked with maximizing a point score, corresponding to a final monetary payment. The task included a two steps decision design. First, subjects choose between exploiting a current option or exploring the environment for an alternative option, paying a travelling cost. When subjects choose to exploit the current option, a second choice between two probabilistic rewards had to be made. This two step design allowed Kolling and colleagues to tell apart the different neural contributions to foraging decision making and typical neuroeconomic binary choice. Their findings indicate that ACC activity reflects the average value of the environment in both stages, and that variability in the exploration choice across subjects correlates with the strength of ACC activity. This stable representation is perfectly suitable to compute foraging decisions. By contrast, the authors found that ventromedial PFC (vmPFC) activity reflects the value and contrast most relevant to the ongoing decision, shifting from *exploration vs exploitation* to *option 1 vs option 2*. This makes vmPFC most suitable to compute value-based (Wunderlich et al., 2012) and multi-attribute comparison decisions (Fellows, 2006).

Using optogenetic to genetically identify parvalbumin (PV) and somatostatin (SOM) inhibitory neurons in the ACC of the mouse, Kvitsiani, Ranade and colleagues (Kvitsiani, Ranade et al., 2013) dug deeper into how ACC can support foraging decisions. In particular, they found that a subset of SOM neurons are selectively active when animals approach a reward, while PV neurons' activity encodes the duration that animals stay in a patch before leaving. Supporting Kolling and colleagues' hypothesis, these results indicate that PV and SOM ACC neurons' activity define the epoch relevant to leaving decisions.

Electrophysiology studies in rats also support Kolling and colleagues' conclusion that vmPFC activity represents the difference in value, changing tracked variables depending on which are most relevant to the current decision. Using a task design that allowed to separate changes due to motor activity from changes due to decision variables, Karlsson and colleagues found that mPFC can play an important role in foraging (Karlsson et al., 2012). Note though that their recordings are located mostly in between the dorsal part of vmPFC and the ventral part of ACC. They identified abrupt changes in the mPFC network that preceded behavioral changes in favour of exploring an alternative option. In addition, they found a transient power increase in high-gamma in the local field potential signalling an upcoming network transition. The authors suggested that such dynamics indicate changes to a state of instability in the network, which gradually settles in a new stable state over the exploration period when new beliefs are established. Similarly, recordings in the mice mPFC have identified neurons whose firing rate changes match modeled changes in relative value, following rewarded or unrewarded outcomes in a two alternatives choice task and this activity is predictive of the following choice (Bari et al., 2019).

Taken together, these results suggest that activity in ACC, is specifically situated to keep track of the environment's average reward and that changes in exploration/exploitation behaviour, which according to MVT depend on the environment richness, are coupled with ACC neural activity (Monosov et al., 2020). Note though that not all authors agree with this interpretation (see Shenav et al., 2014 and 2016; Ebitz and Hayden, 2016).

A second complex computational aspect required in foraging is the ability to constantly update knowledge about the environment. In most of our review of the OFT, we discussed foraging in stable

environments. However, animals need to be able to quickly adjust their behavior in response to different kinds of pressures. For example, a forager could visit a patch that is typically safe to only encounter a dangerous predator. A predator instead, depending on its fitness state, could find a typically manageable prey to be deadly. Therefore, when changes in a situation make the typical behaviour detrimental, the learned values of patches or preys used in OFT need to be flexible and adjustable, possibly quickly. In computational terms, we can describe these changes as transitions between different external/internal states.

For the brain to build representations of such states it requires to combine and constantly update multiple information at the same time. Among the regions forming the PFC, the orbitofrontal cortex (OFC) stands out for the amount of evidence linking it to behavioural adjustments to state changes (see e.g., Roberts, 2006; Izquierdo et al., 2017). A study by Wilson and colleagues have suggested that the OFC is crucial to represent environment states that are not explicitly defined by sensory information, also known as hidden states (Wilson et al., 2014). To investigate the role of OFC in reversal learning (where subjects have to relearn task contingencies after they are reversed unannounced), the authors hypothesized specifically that “OFC provides an abstraction of currently available information in the form of a labeling of the current task state, which is used for reinforcement learning (RL) elsewhere in the brain” (Wilson et al., 2014).

Briefly, RL is a “computational approach to learning from interaction” (Sutton and Barto, 1998) that has successfully helped explain a variety of findings including classical conditioning behavior (Rescorla and Wagner, 1972), and neural activity of dopaminergic midbrain neurons (Schultz et al., 1997). Central to RL is the concept of “state representation”, an abstract and categorical variable that is used

to parse the environment in a set of possible conditions. RL employs a set of algorithms to assign values to each of these states and/or to each possible action performable in a given state. The difference between a poor and an operative state representation is the difference between failure and success, or death and survival in foraging. Since in reversal learning the states are not cued to the animal but have to be learned and recognized by combining a multitude of information about different aspects of the environment, such an operation likely requires executive functions. In this context, Wilson and colleagues specifically suggest that while state representations can exist in various areas of the brain, OFC is uniquely necessary to form conceptual distinctions between states that present perceptually similar features. Thus animals lacking OFC could still solve basic tasks using only observable phenomena and slowly reinforcing the most rewarding action, but cannot use conceptual state to quickly shift strategy. In their work, Wilson et al. show how modelling the difference between control and OFC lesioned animals with a model capable of representing conceptually different states or not, respectively, can recapitulate the effect of the lesion. Their model explains results applicable to a variety of species on a variety of tasks known to engage OFC such as reversal tasks (Teitelbaum, 1964; Butter, 1969; Thorpe et al., 1983; Dias et al., 1996; Schoenbaum et al., 2002; Mc Alon and Brown, 2003; Chudasama and Robbins, 2003; Reekie et al., 2008; Izquierdo et al., 2017), delayed alternation (Mishkin et al., 1969; Miller and Orbach, 1972; Butters et al., 1973), extinction (Butter et al., 1963-1969; Itami and Uno, 2002; Happaney and Zelazo, 2004), and devaluation (Gallagher et al., 1999; Pickens et al., 2003; Izquierdo et al., 2004). Altogether these results indicate that OFC plays a central role in adapting behavior to external (stimulus/action-reward coupling) and internal (reward extinction and devaluation) changes, and that this function could be ascribed to the ability of using conceptual state

representations. However, the employment of different techniques to the study of OFC's role in reversal learning has resulted in diverse behavioral effects, indicating that a more fine grain distinction is required (Rudebeck et al., 2013; Rudebeck and Izquierdo, 2021).

In conclusion, it is evident that while the simplicity of the OFT offers an extremely effective and versatile computational framework to explain behavior, in order to study how the brain solves this challenge, we need to enter in greater details of the decision process itself. Such details often regard how information is combined in order to create operational variables (e.g., long term averages or conceptual variables) that can be used to optimise foraging to an always changing situation. These questions are not only providing further understanding of foraging, but in addition, are helping us to tell apart the contributions of different brain areas, among which the PFC plays a central role.

In Chapter 4 we interfered with PFC areas involved in representing such higher level quantities while mice performed a foraging task. We found that inactivating such areas can lead to drastic qualitative changes in foraging behavior. Our results support the centrality of executive functions in the ability to optimise foraging decisions.

References

- Baddeley, A., & Sala, S. Della. (1998). Working memory and executive control. In A. C. Roberts, T. W. Roberts, & L. Weiskrantz (Eds.), *The Prefrontal Cortex: Executive and Cognitive Functions* (pp. 9–21). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780198524410.003.0002>
- Bari, B. A., Grossman, C. D., Lubin, E. E., Rajagopalan, A. E., Cressy, J. I., & Cohen, J. Y. (2019). Stable Representations of Decision Variables for Flexible Behavior. *Neuron*, 103(5), 922-933.e7.
<https://doi.org/10.1016/j.neuron.2019.06.001>

Bird, D. W., Bird, R. B., & Coddling, B. F. (2009). In pursuit of mobile prey: Martu hunting strategies and archaeofaunal interpretation. *American Antiquity*, 74(1), 3–29. <https://doi.org/10.1017/S000273160004748X>

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbit (Ed.), *Methodology of Frontal and Executive Function* (pp. 81–116). Psychology Press.
<https://www.taylorfrancis.com/chapters/edit/10.4324/9780203344187-8/theory-methodology-executive-function-research-paul-burgess>

Butter, C. M. (1969). Perseveration in extinction and in discrimination reversal tasks following selective frontal ablations in *Macaca mulatta*. *Physiology and Behavior*, 4(2), 163–171. [https://doi.org/10.1016/0031-9384\(69\)90075-4](https://doi.org/10.1016/0031-9384(69)90075-4)

Butter, C. M., Mishkin, M., & Rosvold, H. E. (1963). Conditioning and Extinction of a Food-Rewarded Response after Selective Ablations of Frontal Cortex in Rhesus Monkeys.

Butters, N., Butter, C., Rosen, J., & Stein, D. (1973). Behavioral effects of sequential and one-stage ablations of orbital prefrontal cortex in the monkey. *Experimental Neurology*, 39(2), 204–214.
[https://doi.org/10.1016/0014-4886\(73\)90223-9](https://doi.org/10.1016/0014-4886(73)90223-9)

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, 9(2), 129–136.
[https://doi.org/10.1016/0040-5809\(76\)90040-X](https://doi.org/10.1016/0040-5809(76)90040-X)

Chudasama, Y., & Robbins, T. W. (2003). Dissociable Contributions of the Orbitofrontal and Infralimbic Cortex to Pavlovian Autoshaping and Discrimination Reversal Learning: Further Evidence for the Functional Heterogeneity of the Rodent Frontal Cortex. *The Journal of Neuroscience*, 23(25), 8771–8780. <https://doi.org/10.1523/JNEUROSCI.23-25-08771.2003>

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* 1996 380:6569, 380(6569), 69–72. <https://doi.org/10.1038/380069a0>

Ebitz, R. B., & Hayden, B. Y. (2016). Dorsal anterior cingulate: a Rorschach test for cognitive neuroscience. *Nature Neuroscience* 2016 19:10, 19(10), 1278–1279. <https://doi.org/10.1038/nn.4387>

Fellows, L. K. (2006). Deciding how to decide: Ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. *Brain*, 129(4), 944–952. <https://doi.org/10.1093/brain/awl017>

Funahashi, S., & Andreau, J. M. (2013). Prefrontal cortex and neural mechanisms of executive function. *Journal of Physiology Paris*, 107(6), 471–482. <https://doi.org/10.1016/j.jphysparis.2013.05.001>

- Gallagher, M., McMahan, R. W., & Schoenbaum, G. (1999). Orbitofrontal Cortex and Representation of Incentive Value in Associative Learning. *Journal of Neuroscience*, 19(15), 6610–6614.
<https://doi.org/10.1523/JNEUROSCI.19-15-06610.1999>
- Happaney, K., & Zelazo, P. D. (2004). Resistance to extinction: A measure of orbitofrontal function suitable for children? *Brain and Cognition*, 55(1), 171–184.
[https://doi.org/10.1016/S0278-2626\(03\)00270-7](https://doi.org/10.1016/S0278-2626(03)00270-7)
- Hayden, B. Y., & Walton, M. E. (2014). Neuroscience of foraging. *Frontiers in Neuroscience*, 8(3), 358–378. <https://doi.org/10.3389/fnins.2014.00081>
- Itami, S., & Uno, H. (2002). Orbitofrontal cortex dysfunction in attention-deficit hyperactivity disorder revealed by reversal and extinction tasks. *NeuroReport*, 13(18), 2453–2457. <https://doi.org/10.1097/00001756-200212200-00016>
- Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, 345(March), 12–26.
<https://doi.org/10.1016/j.neuroscience.2016.03.021>
- Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral Orbital Prefrontal Cortex Lesions in Rhesus Monkeys Disrupt Choices Guided by Both Reward Value and Reward Contingency. *Journal of Neuroscience*, 24(34), 7540–7548.
<https://doi.org/10.1523/JNEUROSCI.1921-04.2004>
- Kamil, A. C., Krebs, J. R. K., & Pulliam, H. R. (1987). *Foraging Behavior* (A. C. Kamil, J. R. Krebs, & H. R. Pulliam (eds.)). Springer US.
<https://doi.org/10.1007/978-1-4613-1839-2>
- Karlsson, M. P., Tervo, D. G. R., & Karpova, A. Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science*, 338(6103), 135–139. <https://doi.org/10.1126/science.1226518>
- Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science*, 336(6077), 95–98.
<https://doi.org/10.1126/science.1216930>
- Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J. Z., & Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature*, 498(7454), 363–366.
<https://doi.org/10.1038/nature12176>
- Mars, R. B., Jbabdi, S., & Rushworth, M. F. S. (2021). A Common Space Approach to Comparative Neuroscience.
<https://doi.org/10.1146/Annurev-Neuro-100220-025942>, 44, 69–86.
<https://doi.org/10.1146/ANNUREV-NEURO-100220-025942>

McAlonan, K., & Brown, V. J. (2003). Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behavioural Brain Research*, 146(1–2), 97–103. <https://doi.org/10.1016/j.bbr.2003.09.019>

Miller, M. H., & Orbach, J. (1972). Retention of spatial alternation following frontal lobe resections in stump-tailed macaques. *Neuropsychologia*, 10(3), 291–298. [https://doi.org/10.1016/0028-3932\(72\)90020-6](https://doi.org/10.1016/0028-3932(72)90020-6)

Mishkin, M., Vest, B., Waxler, M., & Rosvold, H. E. (1969). A re-examination of the effects of frontal lesions on object alternation. *Neuropsychologia*, 7(4), 357–363. [https://doi.org/10.1016/0028-3932\(69\)90060-8](https://doi.org/10.1016/0028-3932(69)90060-8)

Miyake, A., & Shah, P. (1999). Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. In A. Miyake & P. Shah (Eds.), *Models of Working Memory*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139174909>

Miyake, A., & Shah, P. (1999). Toward Unified Theories of Working Memory: Emerging General Consensus, Unresolved Theoretical Issues, and Future Research Directions. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 442–482). Cambridge University Press. <https://doi.org/10.1017/cbo9781139174909.016>

Monosov, I. E., Haber, S. N., Leuthardt, E. C., & Jezzini, A. (2020). Anterior Cingulate Cortex and the Control of Dynamic Behavior in Primates. *Current Biology*, 30(23), R1442–R1454. <https://doi.org/10.1016/j.cub.2020.10.009>

Parker, G. A. (1991). Marginal Value Theorem with Exploitation Time Costs: Diet, Sperm Reserves, and Optimal Copula Duration in Dung Flies.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, 3(9), 337–344. [https://doi.org/10.1016/S1364-6613\(99\)01362-5](https://doi.org/10.1016/S1364-6613(99)01362-5)

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C., & Schoenbaum, G. (2003). Different Roles for Orbitofrontal Cortex and Basolateral Amygdala in a Reinforcer Devaluation Task. *Journal of Neuroscience*, 23(35), 11078–11084. <https://doi.org/10.1523/JNEUROSCI.23-35-11078.2003>

Pineda, D., Ardila, A., Rosselli, M., Cadavid, C., Mancheno, S., & Mejia, S. (1998). Executive Dysfunctions in Children with Attention Deficit Hyperactivity Disorder. *International Journal of Neuroscience*, 96(3–4), 177–196. <https://doi.org/10.3109/00207459808986466>

Preuss, T. M. (1995). Do rats have prefrontal cortex? The Rose-Woolsey-Akert program reconsidered. *Journal of Cognitive Neuroscience*, 7(1), 1–24. <https://doi.org/10.1162/jocn.1995.7.1.1>

Pyke, G. H., Pulliam, H. R., & Charnov, E. L. (1977). OPTIMAL FORAGING : A SELECTIVE REVIEW OF THEORY AND TESTS. *The Quarterly Review of Biology*, 52(2), 137–154. <https://doi.org/10.1086/409852>

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of Frontal and Executive Function* (pp. 1–37). Psychology Press.
<https://www.taylorfrancis.com/chapters/edit/10.4324/9780203344187-5/introduction-methodologies-models-study-executive-function-patrick-rabbitt>

Reekie, Y. L., Braesicke, K., Man, M. S., & Roberts, A. C. (2008). Uncoupling of behavioral and autonomic responses after lesions of the primate orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, 105(28), 9787–9792. <https://doi.org/10.1073/PNAS.0800417105>

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement Human Learning View project Taste Aversion Learning View project. In A.H. Black & W.F. Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory*, March, 64–99. <https://www.researchgate.net/publication/233820243>

ROBERTS, A. (2006). Primate orbitofrontal cortex and adaptive behaviour. *Trends in Cognitive Sciences*, 10(2), 83–90.
<https://doi.org/10.1016/j.tics.2005.12.002>

Roberts, A. C. (1998). Introduction. In A. C. Roberts, T. W. Robbins, & L. Weiskrantz (Eds.), *The Prefrontal CortexExecutive and Cognitive Functions* (pp. 1–8). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780198524410.003.0001>

Roberts, A. C., Robbins, T. W., & Weiskrantz, L. (1998). The Prefrontal CortexExecutive and Cognitive Functions. In A. C. Roberts, T. W. Robbins, & L. Weiskrantz (Eds.), *The Prefrontal CortexExecutive and Cognitive Functions*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780198524410.001.0001>

Rudebeck, P. H., Saunders, R. C., Prescott, A. T., Chau, L. S., & Murray, E. A. (2013). Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nature Neuroscience*, 16(8), 1140–1145.
<https://doi.org/10.1038/nn.3440>

Rudebeck, P. H., & Izquierdo, A. (2021). Foraging with the frontal cortex: A cross-species evaluation of reward-guided behavior. *Neuropsychopharmacology*, April. <https://doi.org/10.1038/s41386-021-01140-0>

Schoenbaum, G., Nugent, S. L., Saddoris, M. P., & Setlow, B. (2002). Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *NeuroReport*, 13(6), 885–890.
<https://doi.org/10.1097/00001756-200205070-00030>

- Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275(5306), 1593–1599. <https://doi.org/10.1126/SCIENCE.275.5306.1593>
- Shallice, T., & Burgess, P. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 125–138). Oxford University Press. <https://psycnet.apa.org/record/1992-97203-006>
- Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, 17(9), 1249–1254. <https://doi.org/10.1038/nn.3771>
- Shenhav, A., Straccia, M. A., Botvinick, M. M., & Cohen, J. D. (2016). Dorsal anterior cingulate and ventromedial prefrontal cortex have inverse roles in both foraging and economic choice. *Cognitive, Affective and Behavioral Neuroscience*, 16(6), 1127–1139. <https://doi.org/10.3758/s13415-016-0458-8>
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283(5408), 1657–1661. <https://doi.org/10.1126/science.283.5408.1657>
- Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. In *Decision Theory Models for Applications in Artificial Intelligence: Concepts and Solutions*. <https://doi.org/10.4018/978-1-60960-165-2.ch004>
- Teitelbaum, H. (1964). A comparison of effects of orbitofrontal and hippocampal lesions upon discrimination learning and reversal in the cat. *Experimental Neurology*, 9(6), 452–462. [https://doi.org/10.1016/0014-4886\(64\)90053-6](https://doi.org/10.1016/0014-4886(64)90053-6)
- Thorpe, S. J., Rolls, E. T., & Maddison, S. (1983). The orbitofrontal cortex: Neuronal activity in the behaving monkey. *Experimental Brain Research* 1983 49:1, 49(1), 93–115. <https://doi.org/10.1007/BF00235545>
- Uylings, H. B. M., & G. van Eden, C. (1991). Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Progress in Brain Research* Vol 85, 85(1940), 31–62.
- Uylings, H. B. M., Groenewegen, H. J., & Kolb, B. (2003). Do rats have a prefrontal cortex? *Behavioural Brain Research*, 146(1–2), 3–17. <https://doi.org/10.1016/j.bbr.2003.09.028>
- Vandeviver, C., Neirynck, E., & Bernasco, W. (2021). The foraging perspective in criminology: A review of research literature. *European Journal of Criminology*. <https://doi.org/10.1177/14773708211025864>
- Wallis, C. U., Cardinal, R. N., Alexander, L., Roberts, A. C., & Clarke, H. F. (2017). Opposing roles of primate areas 25 and 32 and their putative rodent homologs in the regulation of negative emotion. *Proceedings of the National*

Academy of Sciences of the United States of America, 114(20), E4075–E4085.
<https://doi.org/10.1073/pnas.1620115114>

Walters, A., Ramiah, V., & Moosa, I. (2016). Ecology and finance: A quest for congruency. *Journal of Behavioral and Experimental Finance*, 10, 54–62.
<https://doi.org/10.1016/J.JBEF.2016.03.006>

Werner, E. E., & Hall, D. J. (1974). Optimal Foraging and the Size Selection of Prey by the Bluegill Sunfish (*Lepomis Macrochirus*). *Ecology*, 55(5), 1042–1052.
<https://doi.org/10.2307/194035>

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, 81(2), 267–279.
<https://doi.org/10.1016/j.neuron.2013.11.005>

Wise, S. P. (2008). Forward frontal fields: phylogeny and fundamental function. *Trends in Neurosciences*, 31(12), 599–608.
<https://doi.org/10.1016/j.tins.2008.08.008>

Wunderlich, K., Dayan, P., & Dolan, R. J. (2012). Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience*, 15(5), 786–791. <https://doi.org/10.1038/nn.3068>

Persistence and flexibility, and their modulation by serotonin

So far, we have considered the key information that should guide foraging decisions (Optimal Foraging Theory) and which brain areas represent such information (executive functions and prefrontal cortex). This section will focus on behavioral tendencies and their neural substrates that can introduce variability and tilt decision-making behavior, particularly towards staying or switching. An important distinction from the previous sections is that here we are less focused on the information represented and the separate computations determining the decision, and more on modulatory factors that, depending on the context, can tune up or down those computations and the resulting behavior.

When describing foraging previously, we summarized the available decision options using the terms exploration and exploitation. Here we will refer to the tendencies towards each with flexibility and persistence, respectively. In addition, we will attempt to bridge disparate sets of literature that we suspect may be related and investigate the same underlying processes. One crucial aspect of reviewing studies on this subject is considering the research framing. The terms flexibility and persistence can be used to refer to diverse aspects (psychological or computational), carrying fundamentally different connotations in distinct tasks. In this chapter, we adopt definitions based on the behavioral advantage provided by each of them.

On the one hand, we use flexibility to refer to the ability to change behavior when it loses efficacy. On the other hand, we define persistence as the capacity to maintain a favorable response in the face of spurious negative outcomes. By adopting these definitions, we want to highlight that the context entirely determines the valence of these

opposing tendencies. This aspect is particularly relevant because most studies we will review here focus on flexibility and persistence in settings where they are maladaptive to the task context (e.g., Clarke et al. 2004, 2007; Lapiz-Bluhm et al., 2009; Bari et al., 2010). In fact, we argue that flexibility and persistence are related to impulsivity and compulsivity, respectively, but are distinguished by them based on whether they are adaptive or maladaptive in a particular task. While it is crucial to focus on the maladaptive side (impulsivity, compulsivity) from a clinical perspective, it is interesting to consider the benefits that these tendencies (flexibility, persistence) may confer when moderately expressed. The latter will be our main focus here.

Before considering neural substrates, it is essential to note that we are taking a high-level approach in drawing the parallel between flexibility/persistence and impulsivity/compulsivity. The impulsivity and compulsivity literature has convincingly demonstrated that these constructs are not monolithic. Indeed, substantial differences in the psychological and neural substrates underpin distinct subtypes. Here, we restricted ourselves to particular subtypes (i.e., waiting impulsivity and cognitive inflexibility), which we will examine in later chapters. For a more detailed discussion on the granularity of these constructs, please refer to the following, more extensive reviews (e.g., Yücel, Fontenelle, 2012; Bari and Robbins, 2013; Figee et al., 2016; Dalley and Robbins, 2017; Van Timmeren et al., 2018; Lüscher, Chamberlain, et al., 2019; Robbins and Crockett, 2020).

Neuromodulators (e.g., serotonin and dopamine) play a central role in our current understanding of persistence and flexibility (Dalley and Robbins, 2017; Robbins and Crockett, 2020; Jones et al., 2021). This fact may appear evident, taking into account two considerations. First, in contrast with executive functions, here we discuss behavioral

tendencies that modulate behavior mainly at a quantitative, not qualitative level. Second, as hinted by the parallel with impulsivity and compulsivity, most existing studies aimed to understand behavioral tendencies relevant to psychiatric disorders, whose treatment heavily relies on drugs acting on neuromodulators. Here we focus on one neuromodulator, in particular, serotonin (5-hydroxytryptamine, 5-HT) which has long been implicated in reward-guided behavior (Fischer and Ullsperger; 2017). It is sufficient to have read a few reviews on this topic to predict that the following sentence of this paragraph would be something along the lines of “but to define its specific role has proven extremely elusive” (Daw et al., 2002; Cools et al., 2008; Luo et al., 2015). This conditioned expectation highlights how central to the field is the quest to define the role of serotonin and how challenging this quest has proven to be. The difficulty in determining a unitary function for 5-HT is also well exemplified by the fact that 5HT is the monoamine involved in treating the most diverse set of psychiatric conditions (Naughton et al., 2000). One primary reason it is hard to define 5-HT’s role is the existence of a large body of apparently contradictory results upon manipulation of the 5-HT system, which we will discuss below. Typically, reviews about 5-HT try to focus on the consistencies and derive a theory that could fit some, even if not all, the data (Soubrié 1986; Daw et al., 2002; Luo et al., 2016). This is a valuable approach in making sense of confusing literature and leaving the reader more informed rather than more confused. Less often, the apparently irreconcilable observations in the literature are at the front and center (e.g., Cools, Roberts, and Robbins, 2008; Rygula et al., 2014). This is undoubtedly helpful to organize the review in ways that would leave the reader more informed rather than confused. However, these apparent discrepancies may be the crucial window into understanding the complex workings of this system. As you will see, our work on 5-HT was deeply inspired by those

discrepancies, and therefore, we chose to structure this section to highlight them.

One promising approach that has helped interpret the paradoxical effects of 5-HT is to consider differences in the environment state (e.g., Seo, Guru, et al., 2019). Seo, Guru, and colleagues have shown that the dorsal raphe 5-HT system produces opposite responses depending on the threatening level of the environment. For instance, stimulation of dorsal raphe nucleus (DRN) 5-HT neurons suppresses or promotes movement in low and high threat environments, respectively.

To evaluate the viability of this context-dependent interpretation to our focus (persistence and flexibility), here, we compare studies that employ the same form of 5-HT manipulation (e.g., global depletion or selective antagonists) but assess its behavioral effects in tasks favoring either flexibility or persistence. One type of settings that prefers flexibility is reversal learning tasks (e.g., Clarke et al., 2004). In these tasks, subjects have to learn associations between stimuli and outcomes. After some time, a reversal in the stimulus-outcome associations (usually swapping the valence associated with the stimuli) occurs covertly. In these tasks, the faster the subjects renounce the previously learned association (i.e., choosing a previously unrewarded option), the better. In contrast, acting on the previous association (i.e., continuing to select the option that is no longer rewarded) is labeled as a perseverative error. In this context (i.e., deterministic reversal learning tasks), the most beneficial strategy is immediately giving up the previously learned associations after encountering one discrepancy.

An example of the opposite case, tasks favoring persistence, are waiting tasks (e.g., Miyazaki et al. 2014). In these paradigms, subjects are required to withhold action for a variable period. If they succeed in

waiting for a randomly delayed signal, they receive a reward. In these designs, every time subjects give up, withholding the action is considered a failure. Thus, while reversal tasks favor switching when any expectation is violated (omission of expected rewards), waiting-type tasks favor persisting despite negative feedback (lack of reward).

Interestingly, similar manipulations of the 5-HT system, either decreasing or increasing 5HT transmission, in these two types of tasks have led to paradoxical results. Several studies in both rodents and primates have now shown that decreasing 5HT transmission via depletion (Lapiz-Bluhm et al., 2009; Bari et al., 2010; Kanen et al., 2021) or inhibition (Matias et al., 2017) while subjects perform a reversal task, results in significantly more perseverative errors. On their own, these results would suggest that 5-HT promotes flexibility. By contrast, in contexts where rewards depend on the ability to persist, in withholding or maintaining an action, global serotonergic depletion (Eagle et al., 2009; Harrison et al., 1997a, 1997b, 1999) or inhibition (Miyazaki et al., 2012; Xu et al., 2017) significantly increases premature leaving, suggesting that 5-HT promotes persistence. This effect appears to be specific to withholding or maintaining actions and not general efforts. Denk and colleagues (2005) have compared the effects of 5-HT depletion, through systemic injection of pCPA (para-chlorophenyl-alanine methyl ester), in two sets of delay discount tasks. In one set, larger rewards were available after climbing over a wall, on the other set after waiting for a 15s delay. Depleting 5-HT selectively reduced larger reward choices in the waiting setting without affecting reward sensitivity (Denk et al., 2005).

The same apparent contradictory effects are reported when 5-HT levels are increased rather than reduced. On the one hand, in reversal tasks, pharmacological enhancement of 5-HT using a selective serotonin reuptake inhibitor (SSRI, Bari et al., 2010; Brigman et al., 2010) reduces

perseverative errors, which suggests an increase in flexibility. On the other hand, in waiting tasks (Miyazaki et al., 2014, 2018; Fonseca et al., 2015; Ohmura et al., 2020) and delay discount tasks (Xu et al., 2017), optogenetic stimulation reduces give-ups, suggesting an increase in animals' persistence. In sum, globally increasing or decreasing 5HT transmission during reversal and waiting-type tasks have paradoxical effects.

Contrary to results obtained with global manipulations of serotonin levels, techniques that act on selective 5-HT receptors appear to produce more coherent results across the two types of tasks. In rats performing a spatial reversal-learning task, systemic 5-HT_{2A} receptor antagonism impairs flexibility, while 5-HT_{2C} receptor antagonism facilitates it (Boulougouris et al., 2008). Interestingly, just like in the reversal tasks, the 5-HT_{2A} and 5-HT_{2C} receptors also have opposing effects in waiting-type tasks. For example, in the 5-choice serial reaction time task, where animals have to wait until a signal is presented to obtain rewards, 5-HT_{2A} antagonists increase persistence, and 5-HT_{2C} antagonists reduce it (Winstanley et al., 2004; Fletcher et al., 2007). If we accept that decreased flexibility is equivalent (or close) to increased persistence, then the effects of each receptor type manipulation are largely consistent. Thus, while global manipulations show opposing effects across tasks, receptor-specific manipulations show consistent results.

The consistency of the effect of these two receptors extends to discrete-trials temporal discrimination tasks, where the best approach is to balance flexibility and persistence rather than maximize one over the other. In these tasks, mice are presented with two levers after a variable interval (Halberstadt et al., 2016). Responding on one lever or the other is reinforced if the interval is shorter or longer than a criterion,

respectively. To contrast these results with other tasks, we could consider choosing the short interval lever akin to selecting the current option (persistence), while choosing the long interval one would be akin to looking for an alternative (flexibility). In this perspective, Halberstadt and colleagues' results are consistent with previous studies: systemic injection of 5-HT_{2A} antagonist increased the time indifference point (persistence in choosing the short-interval lever), whereas 5-HT_{2C} antagonist reduced it (flexibility in selecting the long-interval lever).

Taken together, these results appear to suggest that 5-HT promotes flexibility in reversal tasks, possibly through the 5-HT_{2a} receptor. At the same time, it enhances persistence in tasks requiring withholding or maintaining an action, perhaps through the 5-HT_{2c}.

However, in considering this hypothesis, it is critical to take into account another layer of complexity. Studies selectively manipulating serotonin projections indicate that the role of the 5-HT system could be hard to interpret through global inactivation/activation experiments (e.g., Ren et al. 2018, 2019). These studies proved that there is functional segregation already at the level of the DRN. 5HT neurons have different activity profiles and behavioral effects depending on the innervated area. In particular, Ren and colleagues (2018) have shown that serotonergic neurons of the dorsal raphe nuclei (DRN), projecting to the OFC and central amygdala (CeA), form parallel subsystems and have partly non-overlapping responses and effects. For instance, both DRN-OFC and DRN-CeA projecting 5-HT neurons are activated by reward, but they have opposite responses to aversive stimuli. Both suppress locomotion, but only DRN-CeA neurons enhance anxiety-like behavior, while only DRN-OFC neurons promote active coping in a forced swimming test. This means that opposing behavioral effects could not only arise from

different receptors but rather or also from different projections being engaged.

Therefore, to propose that 5HT paradoxical effects can be explained by the engagement or balance between 5HT_{2A} and 5HT_{2C} receptors, a necessary, but not sufficient, condition is to confirm that opposing effects are elicited when acting on the same projections. In non-human primates, it has been consistently shown that depletion of 5-HT projection to the OFC leads to increased perseverance (reduced flexibility) in reversal learning tasks (Clarke et al., 2004; Roberts, 2011). This effect is selective to reversal as it does not impair attentional set-shifting (Clarke et al., 2005). Moreover, this loss of OFC functionality is selectively dependent on 5-HT, as dopaminergic depletion in OFC does not impair reversal (Clarke et al., 2007). Furthermore, the role of 5HT in reversal is selectively dependent on the projection target. Although reversal learning requires both OFC and the caudate nucleus (Clarke et al., 2008), 5-HT depletion in the OFC, but not in the caudate nucleus, impairs reversal. By contrast, dopamine depletion in the caudate nucleus, but not in the OFC, impairs reversal (Clarke et al., 2011). Results confirming the selectivity of serotonin projection to OFC in reversal tasks have been shown as well in rodents. Rats' extracellular 5-HT levels have been shown to increase in OFC during reversal learning, while chronic intermittent cold (CIC) stress blocks this enhancement and impairs reversal (Lapiz-Bluhm et al., 2009).

The literature exploring OFC 5-HT during persistence-bias tasks is less extensive. Nevertheless, those studies that have investigated it suggest that 5-HT has opposing effects in this other context. For instance, Miyazaki and colleagues (2020) showed that optogenetic stimulation of DRN 5-HT projections to the OFC during a waiting task promotes persistence (rather than reducing it as expected from the

reversal studies). This is in line with previous electrophysiological recordings showing that OFC neuron firing predicts delayed reward, and optogenetic stimulation of DRN 5-HT modulates their firing (Zhou et al., 2015). Thus, the existing literature is consistent with the idea that opposing effects can be elicited when acting on the same projections, and 5-HT_{2a}/5-HT_{2c} opponency may explain the apparent paradox.

The last element to assess the viability of the proposed context-dependent hypothesis is whether 5-HT_{2a} and 5HT_{2c} produce opposing effects when selectively manipulated in OFC. Also, there is not a large body of studies to review in this case, and in addition, the results are not entirely consistent. Boulougouris and Robbins (2010) have shown that 5HT_{2c} antagonist directly infused into the OFC improves reversal, in line with the hypothesis that this receptor could actualize the persistence-promoting effect of 5-HT, while 5-HT_{2a} antagonist had no effect. One possible explanation for this inconsistency is the high level of intrinsic activity of the 5-HT_{2a} receptor, which antagonists might not block, but only inverse agonists (Egan et al., 1998). However, Furr and colleagues (2012) have shown that antagonism of the 5HT_{2a} receptor into the OFC impairs reversal, consistent with the proposition that it actualizes the flexibility-promoting effect of 5-HT. To our knowledge, no study has so far looked into how the activity of 5-HT_{2a} and 5-HT_{2c} in the OFC affects persistence in steady tasks. However, it has been shown that the ratio of 5-HT_{2a}/5-HT_{2c} expression in the mPFC is inversely correlated with persistence (Anastasio et al., 2015). In support of the idea that the 5-HT system function could be about balancing tendencies, Anastasio and colleagues (2015) have shown that artificial downregulation of the 5-HT_{2c} receptor results in upregulation of the 5-HT_{2a} one.

To summarise, the studies reviewed above indicate that i) 5-HT_{2a} and 5HT_{2c} promote flexibility and persistence, respectively, ii) global manipulation of 5-HT promotes flexibility or persistence depending on which is more beneficial to the task requirements, and iii) these opposing effects can be elicited by projection in the same area.

Contrary to the laboratory setting, it is unlikely to encounter black and white situations where only one of these tendencies is beneficial in the natural environment. Behavioral flexibility and persistence likely coexist in a continuum, and the balance between them determines behavior. Therefore we reasoned that a better understanding of the relationship between serotonin and flexibility/persistence would require a setting where an animal must balance both of these tendencies to maximize reward, such as in foraging. We have discussed before how animals have to optimize for how long to exploit a depleting patch or bear a cost to explore an alternative in patch leaving tasks. In this context, enduring attempts on the same patch are an index of persistence, while switching to another patch indicates flexibility. Crucially though, the optimal solution lies in the balance between the two, offering possibly a better perspective on the role of 5-HT. Chapter 3 will present our first results in testing this hypothesis. The result summary presented here attempts to reconcile a large body of literature to formulate a testable hypothesis. However, focusing on the commonalities leaves aside essential differences. For instance, the pharmacology of 5-HT_{2a} and 5-HT_{2c} receptors is complicated by their constitutive activity (Berg et al., 2001, 2005). Also, the 5-HT receptors family is far more extended, and other receptors certainly play a role as well (McGowan and Reynolds, 2020). Finally, although related, the various tasks discussed are known to measure different aspects of compulsivity and impulsivity. This distinction in itself might play a central

role in the effect of 5-HT (Dalley and Robbins, 2017; Robbins and Crocket 2020; Gillan et al., 2016).

References

- Anastasio, N. C., Stutz, S. J., Fink, L. H. L., Swinford-Jackson, S. E., Sears, R. M., Dileone, R. J., ... Cunningham, K. A. (2015). Serotonin (5-HT) 5-HT_{2A} Receptor (5-HT_{2AR}):5-HT_{2CR} Imbalance in Medial Prefrontal Cortex Associates with Motor Impulsivity. *ACS Chemical Neuroscience*, 6(7), 1248–1258. <https://doi.org/10.1021/acscchemneuro.5b00094>
- Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: behavioral and neural basis of response control. *Progress in Neurobiology*, 108, 44–79. <https://doi.org/10.1016/J.PNEUROBIO.2013.06.005>
- Bari, A., Theobald, D. E., Caprioli, D., Mar, A. C., Aidoo-Micah, A., Dalley, J. W., & Robbins, T. W. (2010). Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology*, 35(6), 1290–1301. <https://doi.org/10.1038/npp.2009.233>
- Bari, A., Theobald, D. E., Caprioli, D., Mar, A. C., Aidoo-Micah, A., Dalley, J. W., & Robbins, T. W. (2010). Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology*, 35(6), 1290–1301. <https://doi.org/10.1038/npp.2009.233>
- Berg, K. A., Stout, B. D., Maayani, S., & Clarke, W. P. (2001). Differences in Rapid Desensitization of 5-Hydroxytryptamine 2A. *Pharmacology*, 299(2), 593–602.
- Berg, K. A., Harvey, J. A., Spampinato, U., & Clarke, W. P. (2005). Physiological relevance of constitutive activity of 5-HT_{2A} and 5-HT_{2C} receptors. *Trends in Pharmacological Sciences*, 26(12), 625–630. <https://doi.org/10.1016/j.tips.2005.10.008>
- Boulougouris, V., Glennon, J. C., & Robbins, T. W. (2008). Dissociable effects of selective 5-HT_{2A} and 5-HT_{2C} receptor antagonists on serial spatial reversal learning in rats. *Neuropsychopharmacology*, 33(8), 2007–2019. <https://doi.org/10.1038/sj.npp.1301584>
- Boulougouris, V., & Robbins, T. W. (2010). Enhancement of Spatial Reversal Learning by 5-HT_{2C} Receptor Antagonism Is Neuroanatomically Specific. *Journal of Neuroscience*, 30(3), 930–938. <https://doi.org/10.1523/JNEUROSCI.4312-09.2010>
- Brigman, J. L., Mathur, P., Harvey-White, J., Izquierdo, A., Saksida, L. M., Bussey, T. J., ... Holmes, A. (2010). Pharmacological or genetic inactivation of

the serotonin transporter improves reversal learning in mice. *Cerebral Cortex*, 20(8), 1955–1963. <https://doi.org/10.1093/cercor/bhp266>

Chamberlain, S. R., Tiego, J., Fontenelle, L. F., Hook, R., Parkes, L., Segrave, R., ... Yücel, M. (2019). Fractionation of impulsive and compulsive trans-diagnostic phenotypes and their longitudinal associations. *Australian and New Zealand Journal of Psychiatry*, 53(9), 896–907. <https://doi.org/10.1177/0004867419844325>

Clarke, H. F., Dalley, J. W., Crofts, H. S., Robbins, T. W., & Roberts, A. C. (2004). Cognitive inflexibility after Prefrontal Serotonin Depletion. *Science*, 304(5672), 878–880. <https://doi.org/10.1126/science.1094987>

Clarke, H. F., Walker, S. C., Crofts, H. S., Dalley, J. W., Robbins, T. W., & Roberts, A. C. (2005). Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. *Journal of Neuroscience*, 25(2), 532–538. <https://doi.org/10.1523/JNEUROSCI.3690-04.2005>

Clarke, H. F., Walker, S. C., Dalley, J. W., Robbins, T. W., & Roberts, A. C. (2007). Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. *Cerebral Cortex*, 17(1), 18–27. <https://doi.org/10.1093/cercor/bhj120>

Clarke, H. F., Hill, G. J., Robbins, T. W., & Roberts, A. C. (2011). Dopamine, But Not Serotonin, Regulates Reversal Learning in the Marmoset Caudate Nucleus. *Journal of Neuroscience*, 31(11), 4290–4297. <https://doi.org/10.1523/JNEUROSCI.5066-10.2011>

Clarke, H. F., Robbins, T. W., & Roberts, A. C. (2008). Lesions of the Medial Striatum in Monkeys Produce Perseverative Impairments during Reversal Learning Similar to Those Produced by Lesions of the Orbitofrontal Cortex. *Journal of Neuroscience*, 28(43), 10972–10982. <https://doi.org/10.1523/JNEUROSCI.1521-08.2008>

Cools, R., Roberts, A. C., & Robbins, T. W. (2008). Serotonergic regulation of emotional and behavioural control processes. *Trends in Cognitive Sciences*, 12(1), 31–40. <https://doi.org/10.1016/j.tics.2007.10.011>

Dalley, J. W., & Robbins, T. W. (2017). Fractionating impulsivity : neuropsychiatric implications. *Nature Publishing Group*, 18(3), 158–171. <https://doi.org/10.1038/nrn.2017.8>

Daw, N. D., Kakade, S., & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks*, 15(4–6), 603–616. [https://doi.org/10.1016/S0893-6080\(02\)00052-7](https://doi.org/10.1016/S0893-6080(02)00052-7)

Denk, F., Walton, M. E., Jennings, K. A., Sharp, T., Rushworth, M. F. S., & Bannerman, D. M. (2005). Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology*, 179(3), 587–596. <https://doi.org/10.1007/s00213-004-2059-4>

Eagle, D. M., Lehmann, O., Theobald, D. E. H., Pena, Y., Zakaria, R., Ghosh, R., ... Robbins, T. W. (2009). Serotonin depletion impairs waiting but not stop-signal reaction time in rats: Implications for theories of the role of 5-HT in behavioral inhibition. *Neuropsychopharmacology*, 34(5), 1311–1321. <https://doi.org/10.1038/npp.2008.202>

Egan, C. T., Herrick-Davis, K., & Teitler, M. (1998). Creation of a constitutively activated state of the 5-hydroxytryptamine_{2A} receptor by site-directed mutagenesis: inverse agonist activity of antipsychotic drugs. *The Journal of Pharmacology and Experimental Therapeutics*, 286(1), 85–90. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9655845>

Figee, M., Pattij, T., Willuhn, I., Luigjes, J., van den Brink, W., Goudriaan, A., ... Denys, D. (2016). Compulsivity in obsessive-compulsive disorder and addictions. *European Neuropsychopharmacology*, 26(5), 856–868. <https://doi.org/10.1016/j.euroneuro.2015.12.003>

Fischer, A. G., & Ullsperger, M. (2017). An Update on the Role of Serotonin and its Interplay with Dopamine for Reward. *Frontiers in Human Neuroscience*, 0, 484. <https://doi.org/10.3389/FNHUM.2017.00484>

Fletcher, P. J., Tampakeras, M., Sinyard, J., & Higgins, G. A. (2007). Opposing effects of 5-HT_{2A} and 5-HT_{2C} receptor antagonists in the rat and mouse on premature responding in the five-choice serial reaction time test. *Psychopharmacology*, 195(2), 223–234. <https://doi.org/10.1007/s00213-007-0891-z>

Fonseca, M. S., Murakami, M., & Mainen, Z. F. (2015). Activation of dorsal raphe serotonergic neurons promotes waiting but is not reinforcing. *Current Biology*, 25(3), 306–315. <https://doi.org/10.1016/j.cub.2014.12.002>

Furr, A., Lapiz-Bluhm, M. D., & Morilak, D. A. (2012). 5-HT_{2A}-receptors in the orbitofrontal cortex facilitate reversal learning and contribute to the beneficial cognitive effects of chronic citalopram treatment in rats. *The International Journal of Neuropsychopharmacology / Official Scientific Journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)*, 15(9), 1295. <https://doi.org/10.1017/S1461145711001441>

Gillan, C. M., Robbins, T. W., Sahakian, B. J., van den Heuvel, O. A., & van Wingen, G. (2016). The role of habit in compulsivity. *European Neuropsychopharmacology*, 26(5), 828–840. <https://doi.org/10.1016/J.EURONEURO.2015.12.033>

Halberstadt, A. L., Sindhunata, I. S., Scheffers, K., Flynn, A. D., Sharp, R. F., Geyer, M. A., & Young, J. W. (2016). Effect of 5-HT_{2A} and 5-HT_{2C} receptors on temporal discrimination by mice. *Neuropharmacology*, 107, 364–375. <https://doi.org/10.1016/j.neuropharm.2016.03.038>

Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1999). Central serotonin depletion impairs both the acquisition and performance of a symmetrically

reinforced go/no-go conditional visual discrimination. *Behavioural Brain Research*, 100(1–2), 99–112. [https://doi.org/10.1016/S0166-4328\(98\)00117-X](https://doi.org/10.1016/S0166-4328(98)00117-X)

Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1997). Doubly dissociable effects of median- and dorsal-raphé lesions on the performance of the five-choice serial reaction time test of attention in rats. *Behavioural Brain Research*, 89(1–2), 135–149. [https://doi.org/10.1016/S0166-4328\(97\)00053-3](https://doi.org/10.1016/S0166-4328(97)00053-3)

Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1997). Central 5-HT depletion enhances impulsive responding without affecting the accuracy of attentional performance: Interactions with dopaminergic mechanisms. *Psychopharmacology*, 133(4), 329–342. <https://doi.org/10.1007/s002130050410>

Jones, J. A., Zuhlsdorff, K., & Dalley, J. W. (2021). Neurochemical substrates linked to impulsive and compulsive phenotypes in addiction: A preclinical perspective. *Journal of Neurochemistry*, 157(5), 1525–1546. <https://doi.org/10.1111/JNC.15380>

Kanen, J. W., Apergis-Schoute, A. M., Yellowlees, R., Arntz, F. E., van der Flier, F. E., Price, A., ... Robbins, T. W. (2021). Serotonin depletion impairs both Pavlovian and instrumental reversal learning in healthy humans. *Molecular Psychiatry* 2021, 1–11. <https://doi.org/10.1038/s41380-021-01240-9>

Lapiz-Bluhm, M. D. S., Soto-Piña, A. E., Hensler, J. G., & Morilak, D. A. (2009). Chronic intermittent cold stress and serotonin depletion induce deficits of reversal learning in an attentional set-shifting test in rats. *Psychopharmacology*, 202(1–3), 329–341. <https://doi.org/10.1007/s00213-008-1224-6>

Li, Y., Zhong, W., Wang, D., Feng, Q., Liu, Z., Zhou, J., ... Luo, M. (2016). Serotonin neurons in the dorsal raphe nucleus encode reward signals. *Nature Communications*, 7(1), 10503. <https://doi.org/10.1038/ncomms10503>

Luo, M., Li, Y., & Zhong, W. (2016). Do dorsal raphe 5-HT neurons encode “beneficialness”? *Neurobiology of Learning and Memory*, 135, 40–49. <https://doi.org/10.1016/j.nlm.2016.08.008>

Luo, M., Zhou, J., & Liu, Z. (2015). Reward processing by the dorsal raphe nucleus: 5-HT and beyond. *Learning & Memory*, 22(9), 452–460. <https://doi.org/10.1101/LM.037317.114>

Lüscher, C., Robbins, T. W., & Everitt, B. J. (2020). The transition to compulsion in addiction. *Nature Reviews Neuroscience* 2020 21:5, 21(5), 247–263. <https://doi.org/10.1038/s41583-020-0289-z>

Matias, S., Lottem, E., Dugué, G. P., & Mainen, Z. F. (2017). Activity patterns of serotonin neurons underlying cognitive flexibility. *ELife*, 6, 1–24. <https://doi.org/10.7554/eLife.20552>

McGowan, O. O., & Reynolds, G. P. (2020). Functional pharmacogenetics of serotonin receptors in psychiatric drug action. *Handbook of Behavioral*

Neuroscience (Vol. 31). Elsevier.

<https://doi.org/10.1016/B978-0-444-64125-0.00048-7>

Miyazaki, K. W., Miyazaki, K., & Doya, K. (2012). Activation of Dorsal Raphe Serotonin Neurons Is Necessary for Waiting for Delayed Rewards. *Journal of Neuroscience*, 32(31), 10451–10457.

<https://doi.org/10.1523/JNEUROSCI.0915-12.2012>

Miyazaki, K. K. W., Miyazaki, K. K. W., Tanaka, K. F., Yamanaka, A., Takahashi, A., Tabuchi, S., & Doya, K. (2014). Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. *Current Biology*, 24(17), 2033–2040. <https://doi.org/10.1016/j.cub.2014.07.041>

Miyazaki, K., Miyazaki, K. W., Sivori, G., Yamanaka, A., Tanaka, K. F., & Doya, K. (2020). Serotonergic projections to the orbitofrontal and medial prefrontal cortices differentially modulate waiting for future rewards. *Science Advances*, 6(48), 1–15. <https://doi.org/10.1126/sciadv.abc7246>

Miyazaki, K., Miyazaki, K. W., Yamanaka, A., Tokuda, T., Tanaka, K. F., & Doya, K. (2018). Reward probability and timing uncertainty alter the effect of dorsal raphe serotonin neurons on patience. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-04496-y>

Miyazaki, K., Miyazaki, K. W., & Doya, K. (2012). The role of serotonin in the regulation of patience and impulsivity. *Molecular Neurobiology*, 45(2), 213–224. <https://doi.org/10.1007/s12035-012-8232-6>

Naughton, M., Mulrooney, J. B., & Leonard, B. E. (2000). A review of the role of serotonin receptors in psychiatric disorders. *Human Psychopharmacology*. [https://doi.org/10.1002/1099-1077\(200008\)15:6<397::AID-HUP212>3.0.CO;2-L](https://doi.org/10.1002/1099-1077(200008)15:6<397::AID-HUP212>3.0.CO;2-L)

Ohmura, Y., Tsutsui-Kimura, I., Sasamori, H., Nebuka, M., Nishitani, N., Tanaka, K. F., ... Yoshioka, M. (2020). Different roles of distinct serotonergic pathways in anxiety-like behavior, antidepressant-like, and anti-impulsive effects.

Neuropharmacology, 167(July), 107703.

<https://doi.org/10.1016/j.neuropharm.2019.107703>

Ren, J., Friedmann, D., Xiong, J., Liu, C. D., Ferguson, B. R., Weerakkody, T., ... Luo, L. (2018). Anatomically Defined and Functionally Distinct Dorsal Raphe Serotonin Sub-systems. *Cell*, 175(2), 472–487.e20.

<https://doi.org/10.1016/j.cell.2018.07.043>

Ren, J., Isakova, A., Friedmann, D., Zeng, J., Grutzner, S. M., Pun, A., ... Luo, L. (2019). Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *ELife*, 8, 1–36.

<https://doi.org/10.7554/eLife.49424>

Robbins, T. W., & Crockett, M. J. (2020). The role of central serotonin in impulsivity, compulsivity, and decision-making: comparative studies in

experimental animals and humans. In *Handbook of Behavioral Neuroscience* (Vol. 31, pp. 531–548). <https://doi.org/10.1016/B978-0-444-64125-0.00031-1>

Roberts, A. C. (2011). The importance of serotonin for orbitofrontal function. *Biological Psychiatry*, 69(12), 1185–1191. <https://doi.org/10.1016/j.biopsych.2010.12.037>

Rygula, R., Clarke, H. F., Cardinal, R. N., Cockcroft, G. J., Xia, J., Dalley, J. W., ... Roberts, A. C. (2015). Role of Central Serotonin in Anticipation of Rewarding and Punishing Outcomes: Effects of Selective Amygdala or Orbitofrontal 5-HT Depletion. *Cerebral Cortex* (New York, N.Y. : 1991), 25(9), 3064–3076. <https://doi.org/10.1093/CERCOR/BHU102>

Seo, C., Guru, A., Jin, M., Ito, B., Sleezer, B. J., Ho, Y.-Y., ... Warden, M. R. (2019). Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode. *Science*, 542(February), 538–542. <https://doi.org/10.1126/science.aau8722>

Soubrié, P. (1986). Reconciling the role of central serotonin neurons in human and animal behavior. *Behavioral and Brain Sciences*, 9(02), 319. <https://doi.org/10.1017/S0140525X00022871>

van Timmeren, T., Daams, J. G., van Holst, R. J., & Goudriaan, A. E. (2018). Compulsivity-related neurocognitive performance deficits in gambling disorder: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, 84, 204–217. <https://doi.org/10.1016/J.NEUBIOREV.2017.11.022>

Winstanley, C. A., Theobald, D. E. H., Dalley, J. W., Glennon, J. C., & Robbins, T. W. (2004). 5-HT_{2A} and 5-HT_{2C} receptor antagonists have opposing effects on a measure of impulsivity: Interactions with global 5-HT depletion. *Psychopharmacology*, 176(3–4), 376–385. <https://doi.org/10.1007/s00213-004-1884-9>

Xu, S., Das, G., Hueske, E., & Tonegawa, S. (2017). Dorsal Raphe Serotonergic Neurons Control Intertemporal Choice under Trade-off. *Current Biology*, 27(20), 3111–3119.e3. <https://doi.org/10.1016/j.cub.2017.09.008>

Yucel, M., & Fontenelle, L. F. (2012). Compulsivity as an endophenotype: The search for a hazy moving target. *Addiction*, 107(10), 1735–1736. <https://doi.org/10.1111/J.1360-0443.2012.03663.X>

Zhou, J., Jia, C., Feng, Q., Bao, J., & Luo, M. (2015). Prospective Coding of Dorsal Raphe Reward Signals by the Orbitofrontal Cortex. *Journal of Neuroscience*, 35(6), 2717–2730. <https://doi.org/10.1523/JNEUROSCI.4017-14.2015>

Objectives and organization of the thesis

The overarching goal of this thesis was to deepen our understanding of the neural mechanisms underlying the speed at which decisions are taken. It is possible to encounter challenging choices in everyday life where the odds are so much stacked against a particular course of action that eventually, most people would take the same strategy. Yet, there is remarkable variability in the time that will take a specific person to change their approach, whether this means taking stakes and dedicating themselves to a project or giving up on something they tried to achieve for a long time. The cost of not timing these decisions correctly could be emotionally or economically harmful. Even worse, some psychiatric conditions are characterized by the almost complete inability to make these choices, such as depression and obsessive-compulsive disorder. Here we tried to understand how the brain balances enduring or switching one's course of action.

In Chapter 2, we characterize an experimental setting that can enable the study of this balance in mice to take advantage of the vast neuronal manipulations available in this animal model. We describe a new foraging task with the optimal strategy to balance persistence and flexibility, together with a heuristic solution potentially computationally accessible to the brain. In addition, we show how the serotonin (5-HT) system plays a role in this decision-making process.

In Chapter 3, we further explore the role of the 5-HT system in foraging and offer a novel interpretation of its role that could reconcile contradictory findings. First, we show how its role is more related to balancing persistence and flexibility than promoting one over another. Second, we dig into potential pharmacological mechanisms that could link 5-HT release modalities to exert such balance.

In Chapter 4, we look at the sort of information weighted in the decision process and demonstrate that mice can accumulate evidence regarding the state of the environment even when there are no available direct cues. In addition, we show that prefrontal cortex areas provide differential contributions to the decision-making process.

Chapter 5 ties together previous findings and shows how projections from the prefrontal cortex to the dorsal raphe neurons, the primary source of 5-HT to the forebrain, affect the balance of persistence and flexibility in foraging. In particular, we provide evidence that the postnatal development of these projections is responsible for behavioral changes in adulthood.

Finally, in Chapter 6, we summarise and reflect on the results of the previous chapters, discussing their implications, limitations, and future directions.

Chapter 2: Activation of serotonin neurons promotes active persistence in a probabilistic foraging task.

Eran Lottem, Dhruba Banerjee, Pietro Vertechi, Dario Sarra, Matthijs
oude Lohuis & Zachary F. Mainen

Serotonin (5-HT) is a central neuromodulator implicated in the regulation of many biological processes and is one of the most important targets of psychoactive drugs^{1, 2}. As a unifying concept for 5-HT's manifold effects, Soubrie³ put forward the hypothesis that a major function of 5-HT is to promote behavioral inhibition. Building on work showing that blockade of 5-HT transmission results in continued responses to stimuli that are no longer rewarding^{4–7}, he argued that 5-HT biases decisions in favor of passive over active responding.

More recently, the study of 5-HT and behavioral inhibition has concentrated chiefly on impulse control^{8–10}. One of the most common tasks used to study impulsive behavior is the five-choice serial reaction time task (5-CSRTT)¹¹. Although this task was not specifically designed to measure impulsivity, animals sometimes respond before stimulus presentation, a behavior indicative of impulsivity, and brain-wide 5-HT depletion increases impulsivity in this task¹².

Another line of experiments focuses on animals' ability to wait in order to obtain reward¹³. Neurons in the dorsal raphe nucleus (DRN; the major source of 5-HT to the forebrain) increase their firing rates during reward waiting¹⁴. Moreover, pharmacological inhibition of these neurons promotes premature leaving¹⁴, whereas optogenetic activation of the same neurons promotes patience^{15, 16}.

All of these results are broadly consistent with the Soubrie theory of behavioral inhibition. By reducing the motivation to act, increased 5-HT levels would reduce the rate of premature responses in the 5-CSRTT and increase the time animals can wait to obtain delayed rewards. However, they are also consistent with a different theory in which 5-HT promotes not passivity or the ability to tolerate inaction, but the patience or persistence to carry out an action that is itself costly or unrewarding but helps lead to delayed benefit. Thus successful waiting, in this alternative account, consists not only in suppressing the urge to respond prematurely but also actively carrying out a behavioral alternative to responding. Indeed, children facing the “marshmallow task”, in which they must keep from eating one tasty treat in order to gain a second one later on, often succeed not only by passively

controlling their impulse to eat the available marshmallow but also by actively distracting themselves by performing alternative behaviors¹⁷. So far, in the tasks used to study 5-HT, waiting and passivity coincide by design and so these alternative explanations cannot be readily distinguished. Indeed, DRN neurons fire during some types of movements¹⁸ and DRN 5-HT neurons show increased activity during reward consumption compared to travel¹⁹, consistent with an involvement in active behaviors.

Here we sought to disambiguate whether 5-HT promotes waiting through behavioral inhibition or through persistence in the context of foraging behavior. In natural environments, resources are typically found in patches that become exhausted over time, facing animals with an exploitation/exploration dilemma: they must choose between working within in a given patch and giving up to travel to a different one. Foraging within a patch is itself subject to uncertainty—that is the income from a patch is itself irregular or probabilistic—then exploitation requires patience. But this kind of patience is active rather than passive: in order to obtain rewards animals must continue to seek for food even when food-gathering attempts are sometimes unsuccessful.

There are a number of studies of optimal foraging in rodents in their natural habitats^{20, 21}, and it has been proposed that certain operant behaviors are closely related to foraging behavior²², but the use of foraging behaviors in a laboratory setting remains a relatively underexplored area (but see refs. ^{23, 24}). We therefore developed a novel probabilistic foraging task in which mice nose-poke (forage) at one of two ports and are rewarded for their pokes (foraging attempts) according to a random probability schedule that decays exponentially to zero. Owing to the probabilistic nature of the reward schedule, in each trial, a mouse will experience rewarded pokes interspersed with unrewarded pokes. We reasoned that if 5-HT promotes behavioral inhibition and passive behavior, then it should suppress poking, resulting in fewer pokes at each port visit. If, instead, 5-HT promotes persistence, it should promote the active behavior, nose-poking, which is required to obtain rewards.

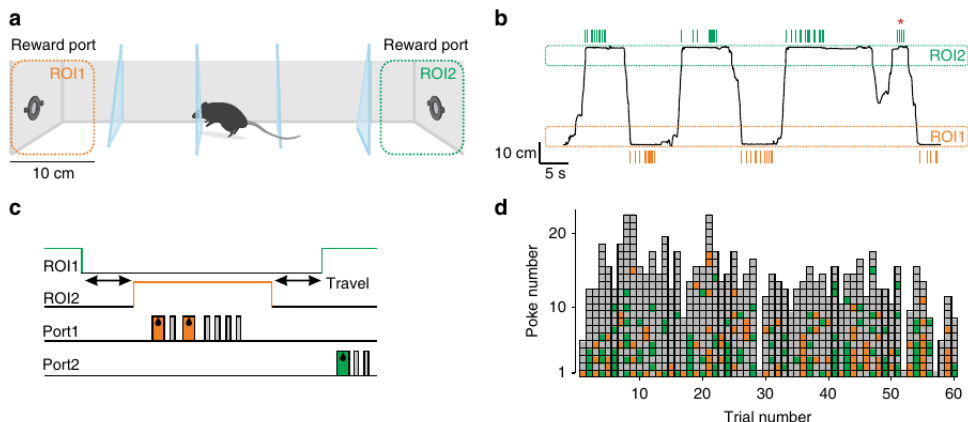


Fig. 1 The probabilistic foraging task. **a** Schematic drawing of the foraging task apparatus. Mice shuttle back and forth between two reward sites, located at the opposite ends of an elongated box, to obtain water rewards. **b** Example snapshot of foraging behavior. The one-dimensional location of an example mouse along the long axis of the box is plotted as a function of time. The ROIs around each water port are marked as dashed rectangles, and green and orange ticks above and below the trajectory mark nose-pokes into the right and left ports, respectively. The red asterisk marks an error trial. **c** Task events during a single trial. Each trial starts with an exit from one of the ROIs. Following shuttling to the other end, the mice would nose-poke multiple times and receive reward on some of the attempts on a probabilistic basis, before switching back. Green and orange rectangles mark rewards, gray rectangles mark omissions. **d** Example session, this time showing only the sequence of outcomes during nose-poking. Each column represents a single trial. Green/orange squares represent rewarded nose-pokes, gray squares, omissions

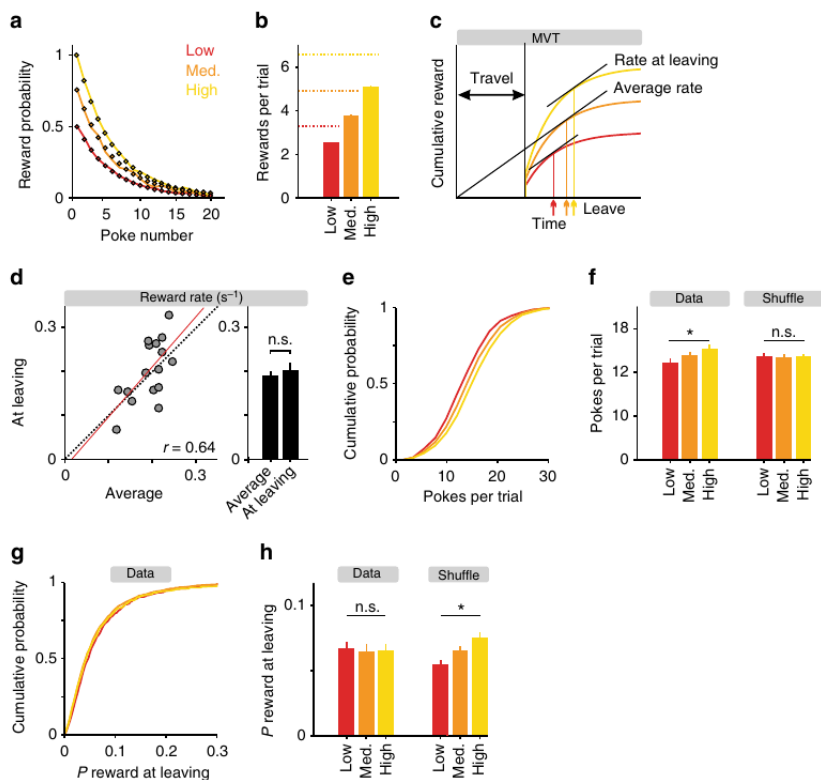


Fig. 2 Reward statistics and task performance. **a** In each trial, reward probabilities were drawn from one of the three exponentially decreasing functions, all sharing the same time constant but with a different scaling factor, labeled high, medium, and low. Dots mark hypothetical values; solid lines are averages derived from data. **b** Bar plot showing the average number of rewards in each of the trial types ($n = 16$ mice). Dashed lines mark maximal values (assuming mice stay in place and poke indefinitely). **c** Schematic drawing of an optimal-agent's behavior during foraging. When plotting the average cumulative reward as a function of time from reward site exit, the average reward rate is the slope of the line that connects this curve at the time of leaving with the origin. Thus the slope of this line is maximal when it is tangent to the curve. Consequently, better or worse trials result in later or earlier leaving times, respectively (vertical lines and arrows). **d** Left: Scatter plot of reward rate at leaving vs. average reward rate. Each circle represents one mouse ($n = 16$). Dashed line is the unity diagonal and red line is a linear regression curve, with its correlation coefficient shown as well ($p < 0.001$). Right: Bar plot showing the average reward rate at leaving and the average reward rate. $p > 0.05$, Wilcoxon signed-rank test. **e** Cumulative distributions of the number of pokes per trial for the three trial types, averaged across mice ($n = 16$). **f** Bar plot showing the average number of pokes in each of the trial types. Bars on the left represent real data, and bars on the right represent shuffled data. Asterisk indicates significant effect ($p < 0.05$, ANOVA). **g** Cumulative distributions of the estimated reward probability after the last poke in a trial (i.e., at the time of switching) for the three trial types, averaged across mice ($n = 16$). **h** Bar plot showing the average reward probability after the last poke. Bars on the left represent real data, and bars on the right represent shuffled data. Asterisk indicates significant difference between trial types, $p < 0.05$, ANOVA.

We found that optogenetic activation of DRN 5-HT neurons increases the number of active nose-pokes a mouse would carry out in an attempt to gain water before giving up. These results contradict the behavioral inhibition hypothesis and support the notion that 5-HT promotes waiting by enhancing persistence in the face of uncertainty and delay.

Results

Mouse behavior in a probabilistic foraging task. To study the role of 5-HT in foraging mice, we developed a novel probabilistic foraging task. Water-restricted mice were placed in a rectangular chamber (50 cm long) containing one water port at each end, which acted as foraging sites (Fig. 1a). Body position was determined using video tracking. A foraging "trial" was considered as a visit to one port, defined as the period between entry and exit into

a region of interest (ROI) around each port (Fig. 1b, c). Correct trials were ones in which the mouse alternated between ports, whereas error trials were those in which the mouse left and then re-entered the same ROI. In correct trials, each nose-poke into the port was either rewarded or not according to a defined probability schedule (Fig. 1d). Reward probabilities were reset to their highest value at the start of each trial and declined exponentially with each poke. Error trials were unrewarded.

To incentivize goal-directed behavior, we introduced three trial types with different initial probability of reward (high-, medium-, and low-quality patches), so that mice would benefit by taking into account the actual rewards received in a given trial, as opposed to adopting a fixed, reward-independent strategy. Reward probability decayed exponentially at the same rate regardless of initial probability (Fig. 2a; Eq. 1 in Methods). The

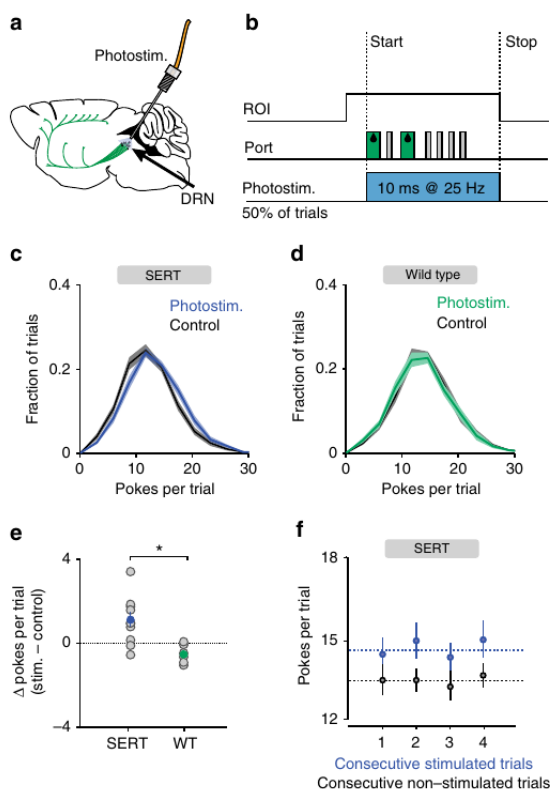


Fig. 3 The effect of DRN 5-HT photostimulation on switching behavior. **a** Scheme of the locations of Chr2-YFP expression and optic fiber placement (adapted with permission from ref. 63). **b** Schematic diagram of task events during a single trial, also showing the period of photostimulation. Photostimulation was triggered by the first poke in 50% of correct trials and ended either when the mouse left the ROI or if 10 s had elapsed since the last poke. **c** Distributions of the number of pokes per trial for photostimulated trials (blue) and control trials (black) averaged across the population of SERT-Cre mice ($n = 10$). **d** Distributions of the number of pokes per trial for photostimulated (green) and control (black) trials, across the population of wild-type mice ($n = 6$). **e** Difference between average number of pokes in photostimulated and control trials for SERT-Cre ($n = 10$) and wild-type ($n = 6$) mice. Averages across mice are shown in filled circles. $*p < 0.05$, Wilcoxon rank-sum test. **f** Plot showing the average number of pokes in photostimulated and control trials for SERT-Cre mice ($n = 10$) as a function of consecutive numbers of photostimulated (blue) or control (black) trials. There was no effect of sequence length on behavior, $p > 0.05$, linear regression analysis

three trial types were presented in a randomly interleaved manner and were not cued. As expected, the average number of rewards gained was highest in high-quality patches and lowest in low-quality patches (Fig. 2b). However, the number of rewards mice obtained was lower than the total available. This could be expected due to a trade-off between continuing to try to exploit the current port and leaving to try the other, at the cost of travel between ports (3.32 ± 1.70 s (mean $\pm$ SD)). This tradeoff is formalized within optimal foraging theory by the marginal value theorem (MVT)²⁵, which states that optimal foragers should leave a depleting resource whenever the instantaneous reward rate

within a patch drops below the long-term average reward rate calculated taking into account travel times (Fig. 2c).

We tested the MVT in our data set and found that the reward rate at the time of leaving was indeed nearly identical to the average reward rate (0.19 ± 0.010 rewards s^{-1} on average and 0.20 ± 0.017 rewards s^{-1} at leaving (mean $\pm$ SEM); $p = 0.54$, Wilcoxon signed-rank test, $n = 16$ mice; Fig. 2d). Two additional predictions of the MVT are that²⁶: (1) the mice should make more pokes in better than average trials, and less in worse than average trials; and (2) reward probability at the time of leaving should be the same regardless of the initial probability at the start of the trial. We tested these predictions in our data set by comparing the number of pokes made (Fig. 2e, f) and the expected reward probabilities at the time of leaving (Fig. 2g, h) for the three trial types. We found that, indeed, the number of pokes increased with increasing initial probability ($F_{(2,45)} = 3.49$, $p = 0.04$, one-way analysis of variance (ANOVA), $n = 16$ mice) and that reward probabilities at the time of leaving were not significantly different across trial types ($F_{(2,45)} = 0.049$, $p = 0.95$, one-way ANOVA, $n = 16$ mice). In contrast, in a shuffled data set in which trials were randomly assigned to one of the three trial types, poke numbers were, as expected, similar across trial types ($F_{(2,45)} = 0.046$, $p = 0.95$, one-way ANOVA, $n = 16$) and reward probabilities at the time of leaving varied with trial quality ($F_{(2,45)} = 4.13$, $p = 0.022$, one-way ANOVA, $n = 16$ mice; see Methods for details of the shuffling procedure). The comparison between shuffled and real data supports the idea that the mice were sensitive to the statistics of individual port visits and suggests that the mice employed a near-optimal strategy in this task.

Activation of DRN neurons prolongs active exploitation. We next examined the effect of activation of DRN 5-HT neurons in the probabilistic foraging task. Of the 16 mice examined, 10 expressed Cre-recombinase under the control of the SERT promoter (SERT-Cre) and 6 were wild-type littermates. Both groups of mice were infected in the DRN with a viral vector containing Cre-dependent channelrhodopsin-2 (AAV2/9-Dio-ChR2-eYFP) and implanted with an optical fiber cannula above the site of infection (Fig. 3a and Supplementary Fig. 2)²⁷. After 9 days of training, we started a 10-day testing period, in which we photostimulated randomly on 50% of the correct trials. Stimulation was triggered by the first nose-poke in each trial and lasted until the end of the trial (Fig. 3b). We found that the number of nose-pokes in photostimulated trials was significantly greater than in control trials in ChR2-expressing mice (14.73 ± 0.55 vs. 13.61 ± 0.53 pokes per trial (mean $\pm$ SEM); $p = 0.014$, Wilcoxon signed-rank test, $n = 10$ mice; see Supplementary Fig. 3 for individual mouse data). This difference was also significant comparing ChR2-expressing and wild-type mice (1.12 ± 0.36 vs. -0.46 ± 0.18 (mean $\pm$ SEM); $p = 0.0047$, Wilcoxon rank-sum test, $n = 10$ SERT-Cre and 6 wild-type mice; Fig. 3c-e). The effect of photostimulation was short-lived, occurring only on photostimulated trials without affecting behavior on subsequent trials (Fig. 3f).

Activation of DRN neurons biases leaving but not travel time. These results contradict the hypothesis that 5-HT promotes behavioral inhibition, as the willingness of mice to perform active nose-poking was enhanced rather than reduced by optogenetic activation of DRN 5-HT neurons. We hypothesized that this phenomenon could reflect a bias in the process underlying the decision of whether to stay and poke again or leave after each nose-poke. Consistent with this idea, we also found that photostimulation, which continued until the mouse exited the ROI, increased the delay to leave the foraging site, i.e., the interval

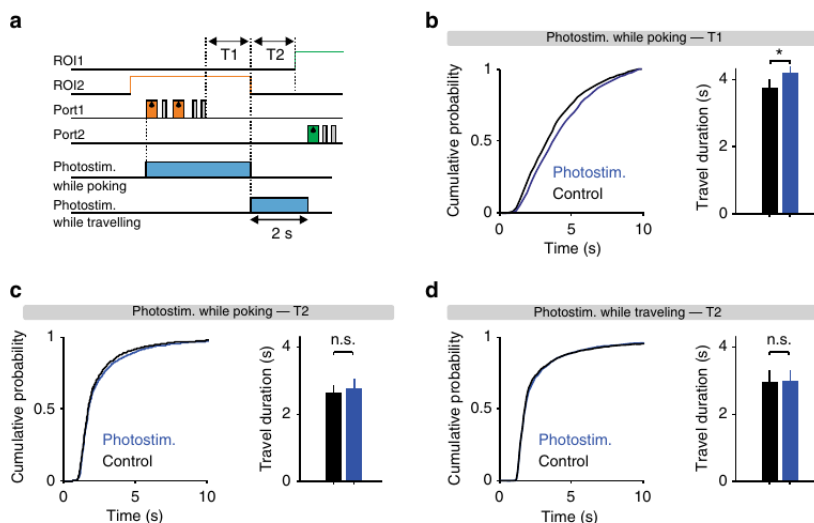


Fig. 4 Lack of effect of DRN 5-HT photostimulation on travel duration. **a** Schematic diagram of task events during a single trial showing the period of photostimulation. In the first protocol (photostimulation while poking), photostimulation was triggered by the first poke in 50% of correct trials, and in the second protocol (photostimulation while traveling), photostimulation was triggered by ROI exit in 50% of the correct trials and lasted for 2 s. We also defined two relevant time periods: (T1) time-to-leave—the interval between the last poke and ROI exit, and (T2) travel—the interval between ROI exit and subsequent ROI entry. **b** Left: Cumulative distributions of T1 durations for photostimulated (blue) and control (black) trials, averaged across the population of SERT-Cre mice ($n = 10$) in the photostimulation-during-poking protocol. Right: Bar plot showing the corresponding averages. $*p < 0.05$, Wilcoxon signed-rank test. **c** Left: Cumulative distributions of T2 durations for photostimulated (blue) and control (black) trials, averaged across the population of SERT-Cre mice ($n = 10$) in the photostimulation-during-poking protocol. Right: Bar plot showing the corresponding averages. $p > 0.05$, Wilcoxon signed-rank test. **d** Left: Cumulative distributions of T2 durations for photostimulated (blue) and control (black) trials, averaged across the population of SERT-Cre mice ($n = 6$) in the photostimulation-during-traveling protocol. Right: Bar plot showing the corresponding averages. $p > 0.05$, Wilcoxon signed-rank test.

between the last poke and ROI exit (control vs. stimulated trials: 3.84 ± 0.25 s in control vs. 4.26 ± 0.23 s in stimulated trials (mean $\pm$ SEM); $p = 0.0039$, Wilcoxon signed-rank test, $n = 10$ mice; Fig. 4b). In contrast, the time it took mice to travel from one port to the other side was not affected (control vs. stimulated trials: 2.64 ± 0.23 vs. 2.77 ± 0.31 s (mean $\pm$ SEM); $p = 0.56$, Wilcoxon signed-rank test, $n = 10$ mice; Fig. 4c). However, the fact that photostimulation was not delivered during the travel period might explain the lack of effect on this measure. Therefore, we subsequently tested 6 of the 10 Chr2-expressing mice from our original cohort using a protocol in which a 2-s stimulation was triggered by ROI exit, so that stimulation took place during the travel period (Fig. 4a). In this case as well, no effect of photostimulation on travel time was observed (control vs. stimulated trials: 3.02 ± 0.31 vs. 3.19 ± 0.33 s (mean $\pm$ SEM); $p = 0.22$, Wilcoxon signed-rank test, $n = 6$ mice; Fig. 4d).

The proportional hazards model of probabilistic foraging. Our analysis thus far suggests that DRN 5-HT neuron stimulation biases the decision process that sets the tradeoff between continued nose-poking ('exploitation') and leaving the site ('exploration'). We next sought to quantify this process more directly. Despite the agreement between our data and the predictions of the MVT (Fig. 2), we deemed it unlikely that the mice were using it directly for two main reasons: (1) the MVT requires an exact calculation of the expected reward probability after each nose-poke, which would require knowledge of the underlying task statistics, memory of the preceding sequence of rewards and omissions, and representation the three reward probabilities

corresponding to the three trial types (Eq. 2 in Methods); (2) the MVT is deterministic in nature and cannot account for the substantial variability we observed in mouse leaving times (this can be seen in the wide range of reward probabilities at the time of leaving in Fig. 2f; the MVT predicts these curves to be step functions)^{28, 29}. Indeed, several simpler heuristics have been proposed to explain decision-making in probabilistic foraging tasks³⁰. Notably, a model known as the "proportional hazards" model³¹ not only aims at predicting mean leaving time but also models a stochastic decision process that leads to leaving^{26, 32–34}. This model assumes that leaving decisions are taken randomly after each omission, based on an underlying hazard function (the conditional probability of leaving after $n + 1$ omissions, having experienced n consecutive omissions) and that the hazard function resets after each reward, albeit to a different value, due to the influence of various predictor variables (Fig. 5a and Eq. 3, see Methods for further details of this model). We compared this model to a stochastic model based on the MVT's predictions (using the expected reward probability as the main predictor of leaving decisions) and found that the proportional hazards provided a better fit to the data in 15 out of the 16 mice tested. The main difference between the two models lies in the resetting property of the proportional hazards model. This feature results in a high win-stay probability, i.e., a high probability of staying after a reward, even for late rewards ("lucky" attempts that do not reflect the actual low reward probability), which was also evident in the data but not in the MVT model (Supplementary Fig. 4).

To test how well the proportional hazards model captures individual mouse behavior, we used half the data to fit model

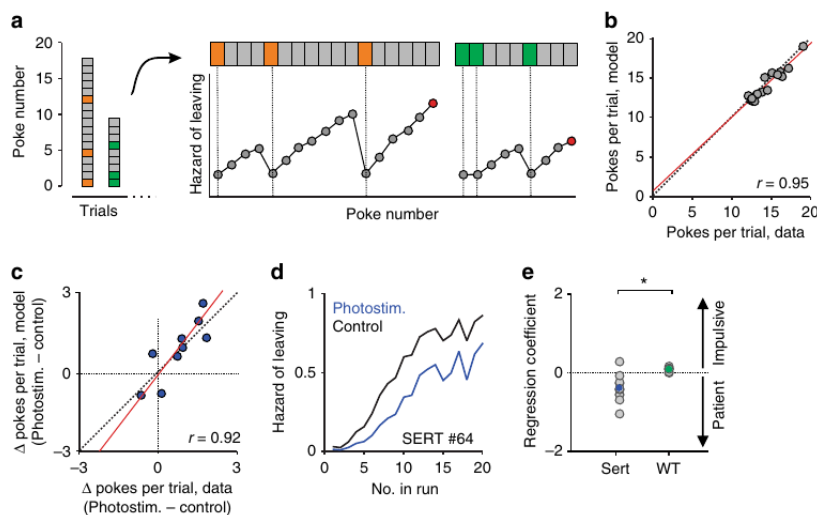


Fig. 5 Fitting behavioral data using the proportional hazards model. **a** Schematic drawing of proportional hazards model fitting pipe-line for two example consecutive trials. Nose-pokes (pooled across trials and sessions for each mouse independently) were used to fit a logistic regression model—the outcome of which was an estimated hazard rate that is reset at trial start and after each reward and is multiplicatively changed by the different coefficients' values. This hazard rate can be viewed as an estimate of the probability of leaving after each nose-poke and can therefore be used to simulate mouse leaving decisions. Leaving decisions depicted as red circles. **b** Scatter plot of simulated vs. real average number of nose-pokes per trial. Each circle represents one mouse ($n = 16$). Dashed line is the unity diagonal and red line is a linear regression curve, with its correlation coefficient shown as well ($p < 0.001$). **c** Same as **b** for the effect of photostimulation on the number of nose-pokes per trial ($p < 0.001$). **d** The modeled hazard function for an example SERT-Cre mouse for photostimulated and control nose-pokes. Note that decreased hazard means longer staying. **e** Cox regression photostimulation coefficient for SERT-Cre ($n = 10$) and wild-type ($n = 6$) mice. Averages across mice are shown in filled circles. * $p < 0.05$, Wilcoxon rank-sum test

parameters and cross-validated it using the second half. These simulations were in excellent agreement with our data (Supplementary Fig. 4). Furthermore, the model was accurate enough to predict idiosyncratic individual differences in poke numbers across mice (Fig. 5b), as well as the effect of photostimulation on a mouse-by-mouse basis (Fig. 5c). When we examined the effect of photostimulation on the estimated hazard rate, expressed as the Cox regression coefficient for that variable, we found it to be significantly negative (that is, divisive or patience promoting) in SERT-Cre mice ($p = 0.014$, Wilcoxon signed-rank test, $n = 10$ mice) and significantly lower than that of wild-type mice ($p = 0.0075$, Wilcoxon rank-sum test, $n = 10$ SERT-Cre and 6 wild-type mice; Fig. 5e).

Next, since the proportional hazards model assumes that stimulation's effect is constant throughout trials (rather than changing as a function of duration), we tested whether this was indeed the case. To address this question, we first recorded multi-unit DRN responses in Chr2-expressing mice to 15 s long trains of optogenetic stimulation. We found that 5-HT neurons kept responding to this stimulation, which was comparable in both frequency and duration to the one used in our behavioral experiments (Supplementary Fig. 5). Since, according to the proportional hazards model, rewards reset the decision process, we reasoned that if the effectiveness of stimulation was changing with time, it should be different for trials in which last reward occurred early vs. late in the trial. However, Supplementary Fig. 6 shows that there was no correlation between the time of the last reward and the effect of photostimulation on the number of subsequent pokes before leaving, suggesting that the effect of stimulation on behavior was constant throughout its duration.

Vigor of nose-poking reports the hazard rate of leaving. The estimated hazard function in the proportional hazards model can be interpreted as a latent decision variable—the instantaneous propensity to switch foraging sites. This variable plays a similar role to that of accumulated evidence in integration-to-bound decision models³⁵. In this interpretation, the effect of optogenetic activation of DRN 5-HT neurons is to slow down the dynamics of the decision variable. To try to test this more directly, we examined the correlation between different behavioral variables related to nose-poking and the latent decision variable in the proportional hazards model (i.e., the estimated hazard of leaving; Supplementary Fig. 7). We found that the duration of nose-pokes for omitted rewards and the duration of inter-poke intervals mirrored the value of the decision variable (Supplementary Fig. 7). Using these two variables, we calculated a measure that we refer to as the “omission duty cycle” (ODC: the ratio between each nose-poke's duration and the sum of the duration and the preceding inter-poke interval; Fig. 6a). The ODC, which essentially measures the rate or vigor of nose-poking, was indeed strongly negatively correlated with the model decision variable (Fig. 6b).

While both the decision variable of the proportional hazards model and the ODC change monotonically through a trial, this pattern is reversed when aligning on the last reward due to the resetting effect of rewards, such that the hazard is higher immediately before the last reward compared to immediately after it (0.056 ± 0.0041 before and 0.014 ± 0.0023 after the last reward (mean $\pm$ SEM); $p = 0.00044$, Wilcoxon signed-rank test, $n = 16$ mice; Fig. 6c, d). We found that, as predicted, this reversal was true for the ODC as well (0.86 ± 0.010 before and 0.87 ± 0.011 after the last reward (mean $\pm$ SEM); $p = 0.0084$, Wilcoxon signed-rank test, $n = 16$ mice; Fig. 6c, d). These results provide evidence

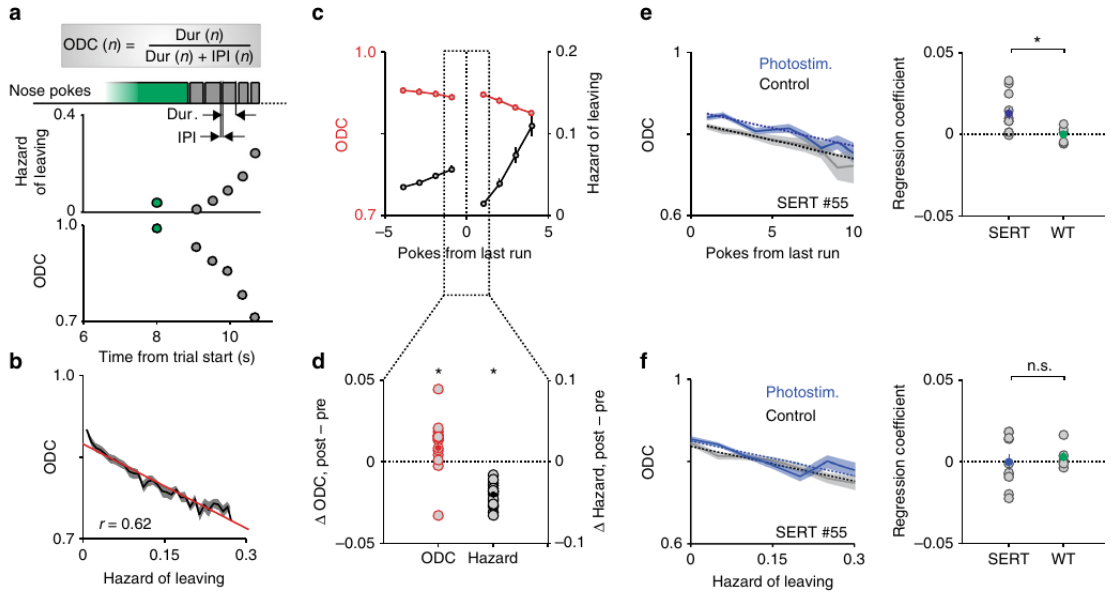


Fig. 6 The effect of DRN 5-HT photostimulation on the microstructure of behavior. **a** Example trial represented in real time and starting from the last reward in that trial. Top: Each rectangle represents a single nose-poke (green marks the last reward and gray the subsequent omissions). Middle: estimated hazard for each nose-poke. Bottom: omission duty cycle (ODC). **b** Correlation between ODC and estimated hazard. The red line is a linear regression curve, with its correlation coefficient shown as well ($p < 0.001$). **c** Rewards reverse an overall (across trial) monotonic decrease in ODC (red) and an overall increase in hazard (black). The plot shows the average ODC and hazard as a function of poke number, aligned on the last reward and averaged across the population of mice ($n = 16$). Note that despite a decreasing (increasing) trend, the ODC (hazard) immediately after the last reward is higher (lower) than the one just before it. **d** Difference between average ODC and hazard immediately before the last reward compared to immediately after it across the population of mice ($n = 16$). $*p < 0.05$, Wilcoxon signed-rank test. **e** Left: The ODC for an example SERT-Cre mouse for photostimulated and control nose-pokes aligned on last reward. The dashed lines are linear regression curves ($p < 0.001$ for the photostimulation coefficient). Right: Regression photostimulation coefficients for SERT-Cre ($n = 10$) and wild-type ($n = 6$) mice. Averages across mice are shown in filled circles. $*p < 0.05$, Wilcoxon rank-sum test. **f** Left: The ODC for the same SERT-Cre mouse shown in **d** for photostimulated and control nose-pokes as a function of estimated hazard. The dashed lines are linear regression curves, ($p > 0.05$ for the photostimulation coefficient). Right: Regression photostimulation coefficients for SERT-Cre ($n = 10$) and wild type ($n = 6$) mice. Averages across mice are shown in filled circles. $p > 0.05$, Wilcoxon rank-sum test

that the vigor with which mice nose-poke is: (1) monotonically decreasing in a manner reflecting the decreasing reward probability; and (2) non-trivially correlated with the effect of rewards on the modeled decision variable. The latter observation argues against the possibility that simpler variables that would be expected to evolve monotonically during nose-poking, such as muscle fatigue, could account for the observed relationships.

Finally, we examined the effect of DRN 5-HT neuron photostimulation on ODC. Since stimulation decreases the hazard of leaving (Fig. 5) and ODC is negatively correlated with the hazard rate, we predicted that the ODC would be higher in photostimulated vs. control trials. Using linear regression analysis, we found that photostimulation indeed increased the ODC of SERT-Cre mice ($p = 0.0098$, Wilcoxon signed-rank test, $n = 10$; see Supplementary Fig. 8 for individual mouse data) and that this increase was significantly higher in SERT-Cre compared to wild-type mice ($p = 0.011$, Wilcoxon rank-sum test, $n = 10$ SERT-Cre and 6 wild-type mice; Fig. 6e). Furthermore, we reasoned that, if the ODC truly reflects the modeled hazard function, then it ought to be the same in photostimulated vs. control trials, when conditioned on the hazard function, since this quantity already takes into account the effect of photostimulation on the decision process. Using similar linear regression analysis, we found that, when plotting the ODC as a function of the modeled hazard, photostimulation had no effect on the ODC of

SERT-Cre mice ($p = 0.90$, Wilcoxon signed-rank test, $n = 10$) nor was there a difference between SERT-Cre compared to wild-type mice ($p = 0.54$, Wilcoxon rank-sum test; Fig. 6f).

These results show that the vigor of poking behavior is closely linked to the underlying decision-making process and can be used as a real-time read-out of a latent decision variable³⁶. Furthermore, the effects of DRN 5-HT neuron stimulation on foraging behavior were well explained by a model in which it biases decisions by multiplicatively scaling the latent decision variable on which leaving decisions are based.

Discussion

In this paper, we found that optogenetic activation of DRN 5-HT neurons during foraging promotes active exploitation of the current foraging site, increasing the time to switch sites. This result in some ways resembles previous findings that DRN 5-HT neuron activation increases the ability to wait for delayed outcomes^{15, 16}. However, because in this task “waiting” required active behavior (nose-poking) the results argue strongly against the hypothesis that 5-HT enhances waiting by enhancing behavioral inhibition. Rather, 5-HT appears to stabilize an on-going behavior by reducing the probability of switching to the next one, i.e., stabilizing nose-poking while reducing the probability of leaving. This behavior could be well described by a simple model,

the proportional hazards model^{26, 32–34}. In this model, the effect of 5-HT stimulation was to proportionately scale down the hazard rate, thereby reducing the probability to leave. Thus 5-HT appears to act as multiplicative bias on the latent decision variable in the foraging decision process.

Foraging behavior is fundamental to any animal's survival. Natural environments require animals to balance various consideration in order to achieve optimal results³⁷. Of particular interest is the exploration–exploitation dilemma: should the animal stay and exploit a depleting resource, or leave instead to forage elsewhere? Within this framework, the MVT describes the optimal strategy, which is deciding to leave at the point in time when the instantaneous rate of rewards drops below its average value²⁵. However, the MVT, as originally formulated, applies to deterministic environments, where foragers have perfect knowledge of the instantaneous and average reward rates. Although these assumptions are not realistic, and the solution for the stochastic case is quite intricate²⁸, a key prediction of optimality remains relevant: switching decisions should be made when instantaneous reward probability is equal to some fixed value, irrespective of whatever has happened up to this point^{26, 33}. Our study confirms that mice follow this prediction.

In order to model the dynamics of the stochastic environment, we used a simple heuristic decision rule, the proportional hazards model³⁴. We found that this model captures not only the fixed relationship between instantaneous and average reward rate at leaving time but also the variability of this decision point across the mouse population (Fig. 5b) and the case-by-case effects of photostimulation across mice (Fig. 5c). Here the effect of DRN 5-HT neuron activation by photostimulation was to decrease the hazard rate in a proportional manner. It thereby increased the average time to reach decision threshold, increasing the time spent nose-poking before deciding to leave.

The proportional hazards model shares commonalities with models that have been widely used in related fields, such as interval timing³⁸ and perceptual decision-making³⁵. In this class of models, before reaching a decision, agents integrate information in the form of elapsed time^{39, 40} or noisy samples of sensory input^{35, 41}. The amount of accumulated information during the accumulation process is a latent or hidden variable that, through behavioral studies, can only be indirectly inferred through modeling the timing of overt behavior: the timing and type of choices made. An important source of evidence in favor of such models has been provided by the ability to find correlates of decision variables in neural activity patterns recorded in brain regions hypothesized to underlie the decision process^{35, 41, 42}. Here we found that a dynamic and readily observable microfeature of mouse behavior, the vigor of nose-poking during foraging, strongly correlated with the latent decision variable in the proportional hazards model. This observation offers support for the validity of the proportional hazards model for foraging. It also supports the notion that readily observable dynamic microfeatures of behavior reflect latent decision variables, providing general insight into the “covert” processes underlying “overt” decisions^{43, 44}.

Our success in modeling the effect of DRN 5-HT neuron stimulation using a simple decision model encourages us that a computationally based and behaviorally constrained approach will be a valuable way to approach the interpretation of 5-HT function. At the same time, our results leave open the question of how more precisely to interpret the influence of 5-HT, since the decision variable that was modulated is highly abstract and represents the conjunction of many more specific factors that might enter into the decision to stay or leave. Broadly speaking, these fall into two categories. First, there are factors that represent the calculation of the costs and benefits of decision options. These

include the cost or effort of foraging at a given site (“handling cost” in the foraging literature), the expected reward gained for each foraging attempt, and the expected cost of travel to the next foraging site. Second, there are factors related to the uncertainties inherent in the probabilistic foraging task, especially the uncertainty related to the stochastic nature of reward delivery³⁷.

For an optimal Bayesian agent, estimating the costs and benefits of decision options must take into account the uncertainties of those estimates. Thus the impact of DRN 5-HT neuron stimulation on the hazard rate of leaving might be due to a more direct effect on costs or benefits of staying vs. leaving or it might arise from a modulation of uncertainty which only indirectly impacts estimates of cost and benefit⁴⁵. We currently favor the latter interpretation for two reasons. First, the endogenous activity of 5-HT neurons reports unexpected events regardless of valence, as appropriate for a signal related to uncertainty⁴⁶. Second, a growing set of evidence suggests that 5-HT does not exert the sort of motivational biases that arise from natural rewards or punishments^{16, 47}. If 5-HT modulated the costs or benefits of on-going events with which it was correlated, one might expect this to be revealed in a broad range of situations. Yet, 5-HT stimulation does not produce place preference or bias thigmotaxis in the open field⁴⁸ and 5-HT stimulation applied during the outcome of a two-alternative choice value-based decision-task does not bias choices¹⁶. While these results are at odds with those reported by Liu et al.⁴⁹, which had found a reinforcing effect of optogenetic DRN 5-HT neuron stimulation on behavior, differences in the transgenic mouse lines used to drive Chr2 expression (ePet1-Cre in Liu et al. vs. SERT-Cre here) or the targeting of different anatomical parts of the DRN may account for this apparent discrepancy⁴⁷. In contrast, elevated 5-HT levels have also been associated with aversive processing^{50, 51}. For example, acute tryptophan depletion (an experimental manipulation used to lower 5-HT levels in humans) attenuates punishment-induced inhibition⁵², suggesting that endogenous 5-HT release may reduce response vigor in the face of aversive predictions⁵³. However, these effects were shown to be specific to aversive contexts, which were absent in the current study. Possible context-dependent differences in 5-HT's effect on foraging behavior could be tested in future experiments by, for example, introducing probabilistic punishments during resource exploitation, and testing the effect of DRN 5-HT stimulation on response vigor under these conditions.

One limitation of the present study is that the task was too complex to be modeled precisely by a simple model; the proportional hazards model is not the optimal solution to this task and the complexity of the optimal solution suggests that it is not likely used by mice. This issue could be addressed in future studies using refined tasks in which simple optimal solutions exist. This, together with additional manipulations of task variables, should help to ascertain more precise interpretation of the influence of 5-HT.

Here we have shown that optogenetic activation of DRN 5-HT neurons prolongs the willingness of mice to forage for stochastically delivered rewards. Waiting can be considered a foraging situation in which rewards are expected but not forthcoming and the same class of integrate-to-threshold models can be used to model both foraging and waiting tasks^{38–40}. On the other hand, in waiting tasks, passivity and waiting are coincident, whereas in foraging tasks, waiting and passivity are decoupled: the choice is no longer between passivity and activity but between two different types of active behavior, nose-poking, or locomoting. The reason that 5-HT stimulation favors patient waiting is apparently not because it favors behavioral inhibition or passivity but because it favors persistence in a current behavior, even if it is active. One possible interpretation of this finding is that 5-HT it

favors persistence by making behavioral transitions (switching from one behavior to another) less likely. This interpretation can explain the observation that DRN photostimulation does not bias travel times between reward sites when stimulation occurs after the mouse is already in transit (Fig. 4 and ref. 48). It would also be consistent with an observed increase in active escape behavior in the forced swim test that is induced by stimulation of medial prefrontal cortex axons in the DRN⁵⁴ and enhanced lose-shift behavior in probabilistic reversal tasks, observed in patients with major depression⁵⁵ and in individuals homozygous for the long 5-HT transporter allele⁵⁶, both associated with decreased levels of extrasynaptic 5-HT. However, as discussed above, another interpretation is that decreased behavioral transitioning is the consequence of a change in one or more underlying decision variables. For example, by increasing the perceived uncertainty of the instantaneous estimate of the reward rate, 5-HT could bias decisions in favor of continued nose-poking in the task⁴⁵.

Despite the tantalizing consistency of the above results, one should not lose sight of the complexity of the 5-HT literature, which implicates this molecule in many functions with little obvious relationship to foraging, waiting, or persistence. However, the application of relatively abstract computational models may provide a useful way toward a more satisfactory overarching understanding. Rather than seeing 5-HT's effects in terms of a suite of diverse behaviors for which the overarching behavioral theme is missing, one may look to commonality of action at the algorithmic level. If 5-HT has a common action on circuits that implement integration-to-bound, then the range of behaviors in which 5-HT acts, such as waiting¹³, cognitive flexibility^{46, 57}, and sensorimotor gain control⁵⁸, will be as diverse as the instances in which this algorithm is applied. For example, in the nematode *Caenorhabditis elegans* it was found that 5-HT favors dwelling in a food patch over roaming⁵⁹, and in larval zebra fish 5-HT is implicated in modulation of short-term locomotor memory to change visuo-motor gain⁶⁰. While the relationship between sensorimotor gain adaptation and probabilistic foraging is obscure at the behavioral level, neural integration is a common component of both. The investigation of 5-HT action at the level of post-synaptic target circuits, together with computational modeling, will be important to test such hypotheses.

Methods

Animal subjects. Mouse lines, surgical procedures for virus injections and optic fiber cannula implantation, and optical set-ups for optogenetic stimulation were identical to those described in previous papers from our laboratory²⁷.

Sixteen adult male C57BL/6 mice (10 SERT-Cre mice and 6 wild-type littermates) were used in this study. All experimental procedures were approved and performed in accordance with the Champsalimaud Centre for the Unknown Ethics Committee guidelines and by the Portuguese Veterinary General Board (Direção-Geral de Veterinária, approval 0421/000/000/2016). The SERT-Cre mouse line⁶¹ was obtained from the Mutant Mouse Regional Resource Centers (stock number: 017260-UCD). The mice were kept under a normal 12 h light/dark cycle, and training as well as testing occurred during the light period. During training and testing the mice were water deprived, and water was available to them only during task performance. Food was freely accessible to the mice in their home cages.

Behavioral training started 2–10 weeks after virus injection and lasted for 9 days, at the end of which we commenced testing (experimentation). Testing periods consisted of 10 consecutive daily sessions. Ten of the 16 mice were only tested once, and 6 SERT-Cre mice (out of the 10) were tested again using a different photostimulation protocol (see below). This second testing period started 1 month after the end of the first one. The experimenters were blind to the mice's genotype throughout training and testing periods.

Adeno-associated virus injection and cannula implantation. Mice were anesthetized with isoflurane (4% induction and 0.5–1% for maintenance) and placed in a stereotaxic frame (David Kopf Instruments, Tujunga, CA). Lidocaine (2%) was injected subcutaneously before incising the scalp. In order to infect serotonergic neurons with ChR2, a craniotomy was drilled over the cerebellum and a pipette

filled with a viral solution (AAV2.9.EF1a.DIO.hChR2(H134R)-eYFP.WPRE.hGH, 10^{13} GC mL⁻¹, University of Pennsylvania) was lowered to the DRN (Bregma –4.7 AP, –2.9 DV) with a 32° angle toward the back of the animal. The viral solution (1 μ L) was injected using a Picospritzer II (Parker). After waiting for 10–15 min, the pipette was removed from the brain and an optical fiber (200 μ m core diameter, 0.48 NA, 4–5 mm long, Doric lenses) was lowered through the same craniotomy such that its tip was positioned 200 μ m above the injection point. The implant was cemented to the skull using dental acrylic (Pi-Ku-Plast HP 36, Bredent, Senden, Germany). Mice were monitored until recovery from the surgery and returned to their home cages. Gentamicin (48760, Sigma-Aldrich, St. Louis, MO) was topically applied around the implant.

Optogenetic stimulation. In order to optically stimulate ChR2-expressing 5-HT neurons, we used blue light from a 473 nm laser (LRS-0473-PFF-00800-03, Laserglow Technologies, Toronto, Canada or DHOM-M-473–200, UltraLasers, Inc., Newmarket, Canada) that was controlled by an acousto-optical modulator (AOM; MTS110-A1-VIS or MTS110-A3-VIS, AA optoelectronic, Orsay, France). Light exiting the AOM was focused into an optical fiber patchcord (200 μ m, 0.22 NA, Doric lenses), connected to a second fiber patchcord through a rotary joint (FRJ 1 $\times$ 1, Doric lenses), which was then connected to the chronically implanted optic fiber cannula.

Histology. In order to confirm successful viral expression of ChR2-eYFP and optical fiber placements, we used postmortem histology at the end of the experiments. Mice were deeply anesthetized with pentobarbital (Eutasil, CEVA Sante Animale, Libourne, France) and perfused transcardially with 4% paraformaldehyde (P6148, Sigma-Aldrich). The brain was removed from the skull, stored in 4% paraformaldehyde overnight, and kept in cryoprotectant solution (PBS in 30% sucrose) for 1 week. Sagittal sections (50 mm) were cut in a cryostat (CM3050S, Leica, Germany) and mounted on glass slides with mowiol mounting medium (81381, Sigma-Aldrich, St. Louis, MO). Scanning images for yellow fluorescent protein (YFP), and transmitted light were acquired with an upright fluorescence microscope (Axio Imager M2, Zeiss, Oberkochen, Germany) equipped with a digital CCD camera (AxioCam MRm, Zeiss) with a 5 $\times$ or 20 $\times$ objective. In a previous study using the same Cre-dependent optogenetic approach and the same mouse line, we reported that 94% of ChR2-eYFP-positive neurons were serotonergic²⁷.

In the analysis shown in Supplementary Fig. 2, we examined the correlation between DRN 5-HT photostimulation and fiber position or ChR2 expression levels. To do so, we first determined, for each mouse, the fiber's distance from the midline by observing which of the sagittal slices contained the most fiber damage. Next, we counted the number of cells in that slice using the ImageJ software's (imagej.net) cell counter plugin.

Electrophysiology. In order to confirm the effectiveness of our photostimulation protocol, we recorded electrophysiological responses in ChR2-eYFP-expressing anesthetized mice (Supplementary Fig. 5). An optrode consisting of an optical fiber (200 μ m diameter) coupled to a 470 nm laser (Laserglow Technologies) and a microelectrode (1–3 M Ω ; FHC) was lowered into the DRN at a 32° angle. Multi-units were acquired digitally using the Spike2 software (Cambridge Electronic Design). Data were stored on a personal computer for offline analysis. Each recording session consisted of 10 15 s long photostimulation sweeps (10 ms pulse width, 25 Hz) and a 60 s interval was inserted between sweeps to allow for activity to return to baseline. In total, we recorded seven multi-units from two ChR2-expressing SERT-Cre mice.

The probabilistic foraging task. Sixteen water-deprived mice were trained in a probabilistic foraging task. The apparatus was an elongated 50 $\times$ 14 cm² chamber with two water-reward ports, one at each end. Each port had an infrared emitter/sensor pair located on its sides to measure nose-pokes (model 007120.0002, Island motion corporation) and a metallic tube running through its center for water delivery (controlled by a valve—LHDA1233115A, The Lee Company, Westbrook, CT). Four partitions were placed at regular intervals along the chamber. Each partition blocked about one half of the corridor's width, forcing the mice to zig-zag between them when crossing from side to side. All task-related events were controlled using a behavioral control system, Bcontrol, that was developed by Carlos Brody (Princeton University) in collaboration with Calin Cuiianu, Tony Zador (Cold Spring Harbor Laboratory), and Z.F.M.

Each trial consisted of a sequence of nose-pokes in one of the two reward ports. Each nose-poke was rewarded with some probability by a 3 μ L drop of water. Reward probabilities decreased after each nose-poke, forcing the mice to alternate between sides during the session. Correct trials were ones in which the mice alternated sides, and error trials were ones in which the mice returned to the same reward port without nose-poking in the other one. Reward probabilities in each correct trial were drawn from one of the three, equally likely exponentials, each

decreasing with poke number

$$P(o_n = 1|t_i) = A_i e^{-\frac{o_n-1}{t_i}} \quad (1)$$

where t_i is the i th trial type ($i = 1, 2, 3$) corresponding to low-, medium-, and high-quality trials. These types differed in their exponential scaling factors, such that $A_1 = 0.5$, $A_2 = 0.75$, $A_3 = 1$. N marks the poke number within a trial, and o_n is the outcome of the n th poke (1 for reward and 0 for omission). Trial types (exponential scaling factors) were randomly interleaved and trial-type identity was not cued to the mice. Reward probability was set to zero during error trials. For technical reasons, the reward probability was set to zero after the 20th nose-poke; this led to only a marginal deviation from true exponentials.

The mice were tracked on-line using Bonsai, a visual programming framework⁶², and their position was used to control trial transitions. We defined a 10 cm square ROI around each port. ROI exits signaled the end of one trial and the beginning of the next one. Additionally, to avoid exceptionally long photostimulation durations, trials were terminated if 10 s had elapsed since the last nose-poke exit, even if the mouse was still in the ROI.

Each testing session lasted about 30 min in which the mice performed at least 80 trials, gaining an average of 3.75 rewards per trial. We found that the behavior of the mice was very consistent throughout the session. However, some trials, particularly towards the end of the session, were unusually short, thus making the distribution of poke numbers per trial bimodal (Supplementary Fig. 1). We therefore consider for analysis only the first 60 trials in each session and only those trials that contained more than two pokes in them.

Photostimulation protocols. In this paper, we present the results of two experiments. The behavioral task was identical in both (as described above), yet the photostimulation protocol differed between the two. In the first experiment, photostimulation (a train of 10 ms wide pulses at 25 Hz, measuring 5 mW at fiber tip) was triggered by the first nose-poke in a trial and lasted until the trial's end (i.e., by either ROI exit, or if 10 s had elapsed since the last nose-poke exit). In the second experiment, a similar photostimulation train was triggered by ROI exit and lasted for 2 s. Sixteen Blue LEDs were placed inside the box, eight above each port, and delivered a flickering masking light, to prevent behavioral changes due to the mere laser light flashes during photostimulated trials. The masking light was identical to the photostimulation in frequency and duration and was present in all trials.

Data analysis. All data analysis was performed using the custom-written software in MATLAB (Mathworks, Natick, MA). When comparing between groups, we used either one-way ANOVA for larger sample sizes ($n = 16$) that met ANOVA assumptions (varstestn and lillietest MATLAB functions testing for homogeneity of variance and normality, respectively) or non-parametric Wilcoxon tests for smaller sample sizes. Linear regression analysis was used to test significant effects of continuous variables. In all figures, average data and error bars, or shaded patches around curves, represent mean $\pm$ SEM.

Subjective reward probability. In the analysis shown in Fig. 2, we calculated the reward probability at the time of leaving (i.e., the expected reward probability in the next poke, if it were to be made). While the actual probabilities are given by Eq. 1, trial types were not cued to the mice, making it impossible to know exactly each poke's reward probability. Instead, we considered the most accurate estimate, assuming perfect knowledge of the task's structure and reward history leading up to a nose-poke. The probability of gaining a reward in poke $n + 1$ given trial history $o_1, \dots, o_n$ is thus:

$$P(o_{n+1} = 1|o_1, \dots, o_n) = \sum_{i=1}^3 P(o_{n+1} = 1|t_i) P(t_i|o_1, \dots, o_n) \quad (2)$$

$$= \sum_{i=1}^3 P(o_{n+1} = 1|t_i) \frac{P(o_1, \dots, o_n|t_i)}{\sum_{j=1}^3 P(o_1, \dots, o_n|t_j)}$$

Where t_i is the i th trial type ($i = 1, 2, 3$) and $P(o_1, \dots, o_n|t_i) = \prod_{j=1}^n P(o_j|t_i)$ is calculated using Eq. 1. In Fig. 2d, we use this quantity to obtain the instantaneous reward rate by calculating, for each nose-poke, the ratio between this and the poke's duration.

Shuffled behavioral data. In Fig. 2, we also compare real to shuffled data sets. Shuffling was done individually for each mouse by first considering all potential trials, namely, the 20 poke long sequences of rewards and misses that are calculated using Eq. 1. Next, we paired each such potential trial with a randomly chosen trial length, selected from the actual data. This resulted in a shuffled data set in which average trial lengths were preserved while all other correlations between trial reward structure (such as trial type) and behavior were destroyed.

Travel time analysis. In Fig. 4, we analyze the effect of photostimulation on travel times. Since the travel time is not well defined in error trials (when the mice leave and return to the same location without visiting the other reward port in-between), we considered for this analysis only those trials in which both current and subsequent trials were correct. Additionally, since photostimulation terminated if 10 s had elapsed since the last nose-poke exit, even if the mouse was still in the ROI, such trials were excluded from this analysis as well. Consistent with the result shown in Fig. 4b, the fraction of such trials was higher in the photostimulation condition (control vs. stimulated trials: $8.30 \pm 2.49\%$ in control vs. $13.30 \pm 3.32\%$ in stimulated trials (mean $\pm$ SEM); $p = 0.002$, Wilcoxon signed-rank test, $n = 10$ mice).

Cox proportional hazards regression model. In order to model individual mouse choice behavior and its modulation by various factors (e.g., DRN photostimulation), we used the Cox proportional hazards regression model. This semi-parametric model calculates the probability of leaving after each poke according to a baseline hazard (the probability of leaving immediately after a poke as a function of its number) that is estimated from the data and that may be changed multiplicatively by trial-general covariates. The model is reset after each reward but potentially to a different value, depending on the covariates. Therefore, we fitted leaving probabilities for all nose-pokes as a function of their distance from the previous reward rather than from trial start. To do so, each trial was segmented into one or more sub-trials or runs. Each such run was bracketed from the left by a reward (except for the first one, in cases where the first poke in a trial was an omission) and from the right (except for the last one, in the very common cases where the last poke in a trial was an omission). Note that length zero runs were also allowed, if two rewards were delivered consecutively. As noted in ref.³¹, in the discrete case this model is equivalent to a logistic model of the form:

$$\frac{\lambda(n)}{1 - \lambda(n)} = e^{\beta x} \frac{\lambda_0(n)}{1 - \lambda_0(n)} \quad (3)$$

where $\lambda(n)$ represents a hazard function (hazard rate of leaving after the n th poke). $\lambda_0(n)$ represents a baseline hazard function, that is, the hazard function when all the predictors are equal to 0. β is a row vector with scalar Cox coefficients for each of the covariates, and x is a column vector representing covariate values. The covariates we used were the position of the previous reward, the port side (this nuisance variable was irrelevant for our interpretation of the data but nonetheless was necessary for accurate fitting), and photostimulation condition (1 in stimulated trials and 0 otherwise).

We also used the fitted model parameters to simulate data. Similarly to the shuffled simulation described above, we first gathered all potential trials from the data. Next, to simulate leaving decisions we used the fitted Cox model to generate a sequence of probabilistic "coin-flips", such that if one assumes "heads" to represent the hazard of leaving after poke n (taking into consideration its distance from the last reward, photostimulation condition, etc.), then for a given trial the simulation would go on with a series of coin-flips until the first "heads" is encountered—this would mark the time of leaving. In these simulations, half of the data were used for fitting the model and the second half, for testing.

To provide further support for our use of the proportional hazards model, we conducted model comparison analysis (Supplementary fig. 4). In it, we compared the proportional hazards model to an alternative model, based on the normative predictions of the MVT and Eq. 2. The model was formulated as follows:

$$\frac{\lambda(n)}{1 - \lambda(n)} = e^{\beta_0 + \beta_1 * P_{rew}(n+1) + \beta_2 * s} \quad (4)$$

where $\lambda(n)$ represents the probability of switching sides after the n th poke. $P_{rew}(n)$ is the n th poke reward probability (calculated using Eq. 2), and s is an indicator variable for side.

Code availability. All the original code used for data analysis is available upon request from Z.F.M. (zmajnen@neuro.fchampalimaud.org).

Data availability. All data are available upon request from Z.F.M. (zmajnen@neuro.fchampalimaud.org).

Received: 21 July 2017 Accepted: 9 February 2018

Published online: 08 March 2018

References

1. Terry, A. V., Buccafusco, J. J. & Wilson, C. Cognitive dysfunction in neuropsychiatric disorders: selected serotonin receptor subtypes as therapeutic targets. *Behav. Brain Res.* **195**, 30–38 (2008).

2. Pytliak, M., Vargová, V., Mechírová, V. & Felšöci, M. Serotonin receptors - from molecular biology to clinical applications. *Physiol. Res.* **60**, 15–25 (2011).
3. Soubrié, P. Reconciling the role of central serotonin neurons in human and animal behavior. *Behav. Brain Sci.* **9**, 319 (1986).
4. Beninger, R. J. & Phillips, A. G. Possible involvement of serotonin in extinction. *Pharmacol. Biochem. Behav.* **10**, 37–41 (1979).
5. Tye, N. C., Iversen, S. D. & Green, A. R. The effects of benzodiazepines and serotonergic manipulations on punished responding. *Neuropharmacology* **18**, 689–695 (1979).
6. Thornton, E. W. & Goudie, A. J. Evidence for the role of serotonin in the inhibition of specific motor responses. *Psychopharmacology (Berl.)* **60**, 73–79 (1978).
7. Graeff, F. G. & Schoenfeld, R. I. Tryptaminergic mechanisms in punished and nonpunished behavior. *J. Pharmacol. Exp. Ther.* **173**, 277–283 (1970).
8. Dalley, J. W. & Robbins, T. W. Fractionating impulsivity: neuropsychiatric implications. *Nat. Rev. Neurosci.* **18**, 158–171 (2017).
9. Basar, K. et al. Nucleus accumbens and impulsivity. *Prog. Neurobiol.* **92**, 533–557 (2010).
10. Cools, R., Roberts, A. C. & Robbins, T. W. Serotonergic regulation of emotional and behavioural control processes. *Trends Cogn. Sci.* **12**, 31–40 (2008).
11. Bari, A., Dalley, J. W. & Robbins, T. W. The application of the 5-choice serial reaction time task for the assessment of visual attentional processes and impulse control in rats. *Nat. Protoc.* **3**, 759–767 (2008).
12. Harrison, A. A., Everitt, B. J. & Robbins, T. W. Central 5-HT depletion enhances impulsive responding without affecting the accuracy of attentional performance: interactions with dopaminergic mechanisms. *Psychopharmacology (Berl.)* **133**, 329–342 (1997).
13. Miyazaki, K., Miyazaki, K. W. & Doya, K. The role of serotonin in the regulation of patience and impulsivity. *Mol. Neurobiol.* **45**, 213–224 (2012).
14. Miyazaki, K., Miyazaki, K. W. & Doya, K. Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. *J. Neurosci.* **31**, 469–479 (2011).
15. Miyazaki, K. W. et al. Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. *Curr. Biol.* **24**, 2033–2040 (2014).
16. Fonseca, M. S., Murakami, M. & Mainen, Z. F. Activation of dorsal raphe serotonergic neurons promotes waiting but is not reinforcing. *Curr. Biol.* **25**, 306–315 (2015).
17. Mischel, W., Ebbsen, E. B. & Raskoff Zeiss, A. Cognitive and attentional mechanisms in delay of gratification. *J. Pers. Soc. Psychol.* **21**, 204–218 (1972).
18. Ranade, S. P. & Mainen, Z. F. Transient firing of dorsal raphe neurons encodes diverse and specific sensory, motor, and reward events. *J. Neurophysiol.* **102**, 3026–3037 (2009).
19. Li, Y. et al. Serotonin neurons in the dorsal raphe nucleus encode reward signals. *Nat. Commun.* **7**, 10503 (2016).
20. Morris, D. W. & Davidson, D. L. Optimally foraging mice match patch use with habitat differences in fitness. *Ecology* **81**, 2061–2066 (2000).
21. Orrock, J. L., Danielson, B. J. & Brinkerhoff, R. J. Rodent foraging is affected by indirect, but not by direct, cues of predation risk. *Behav. Ecol.* **15**, 433–437 (2004).
22. Fantino, E. & Abarca, N. Choice, optimal foraging, and the delay-reduction hypothesis. *Behav. Brain Sci.* **8**, 315–330 (1985).
23. Wikenheiser, A. M., Stephens, D. W. & Redish, A. D. Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task. *Proc. Natl. Acad. Sci.* **110**, 8308–8313 (2013).
24. Kane G.A., Bornstein A.M., Shenhav A., Wilson R.C., Daw N.D. & Cohen J.D. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. (eds, G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar) 637–642 (Austin, TX, Cognitive Science Society, 2017).
25. Charnov, E. L. Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* **9**, 129–136 (1976).
26. Wajnberg, E., Bernhard, P., Hamelin, F. & Boivin, G. Optimal patch time allocation for time-limited foragers. *Behav. Ecol. Sociobiol.* **60**, 1–10 (2006).
27. Dugué, G. P. et al. Optogenetic recruitment of dorsal raphe serotonergic neurons acutely decreases mechanosensory responsivity in behaving mice. *PLoS ONE* **9**, e105941 (2014).
28. Oaten, A. Optimal foraging in patches: a case for stochasticity. *Theor. Popul. Biol.* **85**, 673–690 (1977).
29. McNair, J. N. Optimal giving up times and the marginal value theorem. *Am. Nat.* **119**, 511–529 (1982).
30. Krebs, J. R., Ryan, J. C. & Charnov, E. L. Hunting by expectation or optimal foraging? *Anim. Behav.* **22**, 953–964 (1974).
31. Cox, D. R., Society, S. & Methodological, S. B. Regression models and life-tables. *J. R. Stat. Soc. Ser. B* **34**, 187–220 (1972).
32. Haccou, P. & Hemerik, L. The influence of larval dispersal in the cinnabar moth (*Tyria jacobaeae*) on predation by the red wood ant (*Formica polyctena*): an analysis based on the proportional hazards model. *J. Anim. Ecol.* **54**, 755–769 (1985).
33. Wajnberg, E., Fauvergue, X. & Pons, O. Patch leaving decision rules and the Marginal Value Theorem: an experimental analysis and a simulation model. *Behav. Ecol.* **11**, 577–586 (2000).
34. Tenhumberg, B., Keller, M. A. & Possingham, H. P. Using Cox's proportional hazard models to implement optimal strategies: an example from behavioural ecology 2. Wasp behaviour model. *Behaviour* **33**, 597–607 (2001).
35. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
36. Wang, A. Y., Miura, K. & Uchida, N. The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nat. Neurosci.* **16**, 639–647 (2013).
37. Pyke, G. H. Optimal foraging: a critical review. *Annu. Rev. Ecol. Syst.* **15**, 523–575 (1984).
38. Killeen, P. R. & Fetterman, J. G. A behavioral theory of timing. *Psychol. Rev.* **95**, 274–295 (1988).
39. Kacelnik, A. & Brunner, D. Timing and foraging: Gibbon's scalar expectancy theory and optimal patch exploitation. *Learn. Motiv.* **33**, 177–195 (2002).
40. Simen, P., Vlasov, K. & Papadakis, S. Scale (in)variance in a unified diffusion model of decision making and timing. *Psychol. Rev.* **123**, 151–181 (2016).
41. Brunton, B. W., Botvinick, M. M. & Brody, C. D. Rats and humans can optimally accumulate evidence for decision-making. *Science* **340**, 95–98 (2013).
42. Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
43. de Gee, J. W., Knapen, T. & Donner, T. H. Decision-related pupil dilation reflects upcoming choice and individual bias. *Proc. Natl. Acad. Sci. USA* **111**, E618–E625 (2014).
44. Murphy, P. R., Vandekerckhove, J. & Nieuwenhuis, S. Pupil-linked arousal determines variability in perceptual decision making. *PLoS Comput. Biol.* **10**, e1003854 (2014).
45. Rushworth, M. F. S. & Behrens, T. E. J. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* **11**, 389–397 (2008).
46. Matias, S., Lottem, E., Dugué, G. P. & Mainen, Z. F. Activity patterns of serotonin neurons underlying cognitive flexibility. *Elife* **6**, e20552 (2017).
47. McDevitt, R. A. et al. Serotonergic versus nonserotonergic dorsal raphe projection neurons: differential participation in reward circuitry. *Cell Rep.* **8**, 1857–1869 (2014).
48. Correia, P. A. et al. Transient inhibition and long-term facilitation of locomotion by phasic optogenetic activation of serotonin neurons. *Elife* **6**, e20975 (2017).
49. Liu, Z. et al. Dorsal raphe neurons signal reward through 5-HT and glutamate. *Neuron* **81**, 1360–1374 (2014).
50. Boureau, Y. L. & Dayan, P. Opponency revisited: competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology* **36**, 74–97 (2011).
51. Dayan, P., & Huys, Q. J. M. Serotonin in affective control. *Annu. Rev. Neurosci.* **32**, 95–126 (2009).
52. Crockett, M. J., Clark, L. & Robbins, T. W. Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. *J. Neurosci.* **29**, 11993–11999 (2009).
53. Crockett, M. J., Clark, L., Apergis-Schoute, A. M., Morein-Zamir, S. & Robbins, T. W. Serotonin modulates the effects of pavlovian aversive predictions on response vigor. *Neuropsychopharmacology* **37**, 2244–2252 (2012).
54. Warden, M. R. et al. A prefrontal cortex-brainstem neuronal projection that controls response to behavioural challenge. *Nature* **492**, 428–432 (2012).
55. Murphy, F. C., Michael, A., Robbins, T. W. & Sahakian, B. J. Neuropsychological impairment in patients with major depressive disorder: the effects of feedback on task performance. *Psychol. Med.* **33**, 455–467 (2003).
56. den Ouden, H. E. M. et al. Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* **80**, 1090–1100 (2013).
57. Clarke, H. F. Cognitive inflexibility after prefrontal serotonin depletion. *Science* **304**, 878–880 (2004).
58. Jacobs, B. L. & Fornal, C. A. Serotonin and motor activity. *Curr. Opin. Neurobiol.* **7**, 820–825 (1997).
59. Flavell, S. W. et al. Serotonin and the neuropeptide PDF initiate and extend opposing behavioral states in *C. elegans*. *Cell* **154**, 1023–1035 (2013).
60. Kawashima, T., Zwart, M. F., Yang, C. T., Mensh, B. D. & Ahrens, M. B. The serotonergic system tracks the outcomes of actions to mediate short-term motor learning. *Cell* **167**, 933–946 (2016).
61. Gong, S. et al. Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).

62. Lopes, G. et al. Bonsai: an event-based framework for processing and controlling data streams. *Front. Neuroinform.* **9**, 7 (2015).
63. Paxinos, G. & Franklin, K. *Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates* (Elsevier, USA 2004).

Acknowledgements

We thank Bassam Atallah and Madalena Fonseca for comments on a previous version of the manuscript. We also thank Gil Costa for support with visual the diagram shown in Fig. 1. This work was supported by the European Research Council (Advanced Investigator Grants 250334 and 671251 to Z.F.M.) and Champalimaud Foundation (Z.F.M.).

Author contributions

E.L., D.B., and Z.F.M. designed the experiments. D.B., E.L., P.V., D.S., and M.o.L. conducted the experiments. E.L. and P.V. analyzed the data. E.L. and Z.F.M. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-03438-y>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

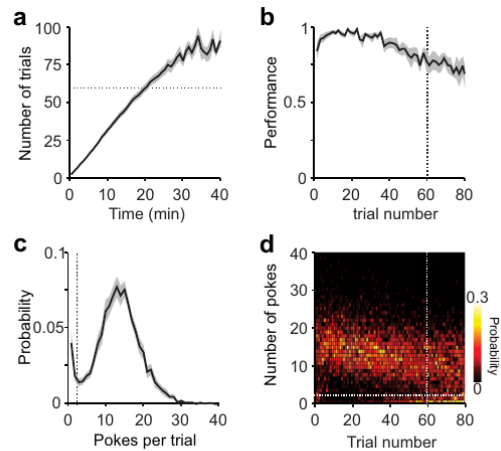
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

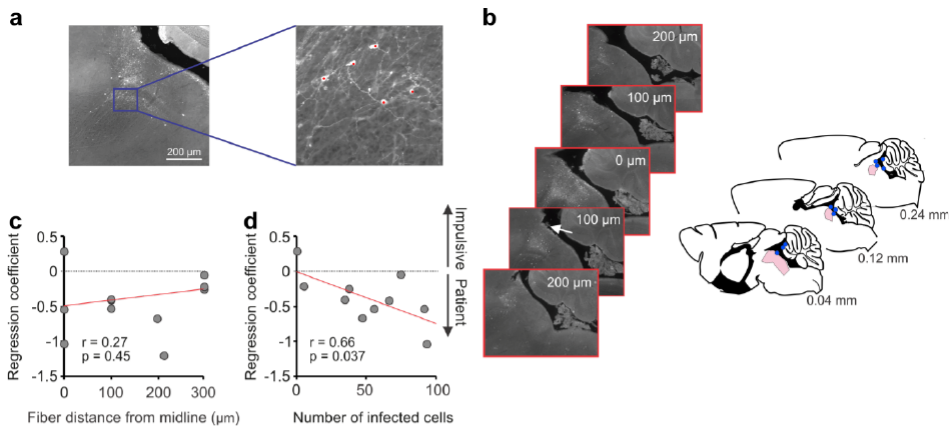
© The Author(s) 2018

Supplementaries



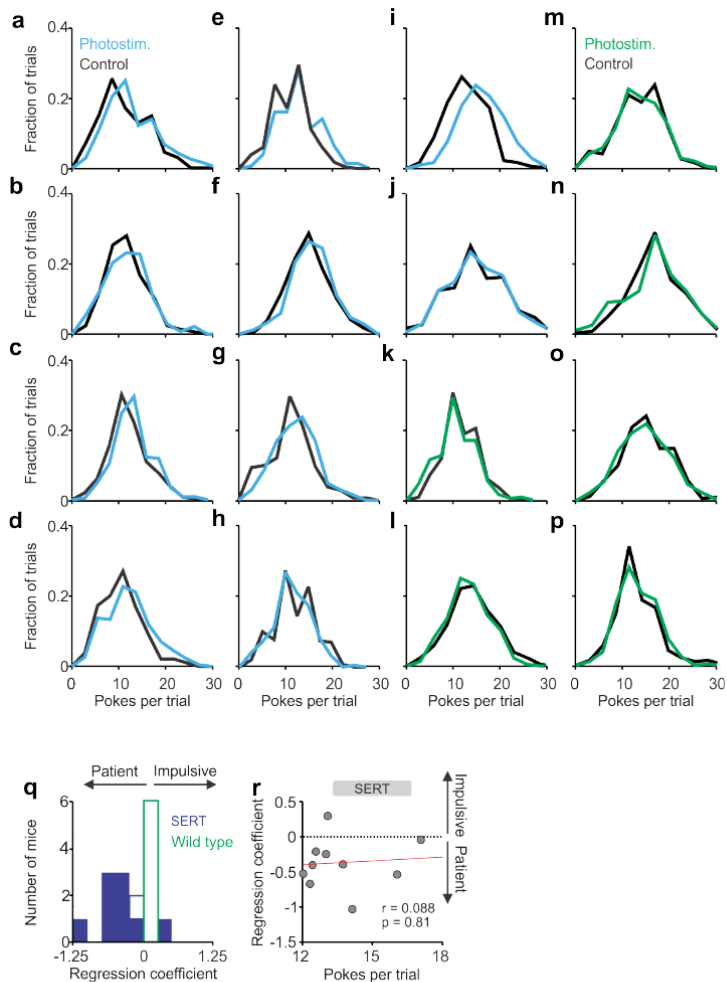
Supplementary figure 1 Task performance and trial selection criteria

- (a) Plot showing average cumulative number of trials as a function of time ($n = 16$). Only trials 1 to 60 were considered for further analysis (cutoff marked by dashed line).
- (b) Plot showing average performance, defined as the probability of correctly alternating between reward sides as a function of trial number (cutoff marked by dashed line).
- (c) Average poke-per-trial distribution ($n = 16$). Note the bimodal shape of the distribution. Trials with less than 3 pokes were considered as “lapses” and discarded from further analysis (dashed line).
- (d) Hit map showing average poke-per-trial distributions as a function of trial number and the two selection criteria (dashed lines marking trial number and number of pokes).



Supplementary figure 2 Histological analysis of virus expression and fiber location

- (a) Left: A fluorescence image of a parasagittal section of dorsal raphe nucleus of an example ChR2-YFP expressing, SERT-Cre mouse. Right: Magnification of the same image, red dots mark labeled neurons.
- (b) Left: Five parasagittal images taken from a different mouse (adapted with permission from Ref. 63). Numbers indicate distance from midline. White arrow points at the location of the fiber tip. Right: Fiber tip locations for the 10 SERT-Cre mice. Dorsal raphe nucleus is indicated in pink.
- (c) Correlation between Cox regression photostimulation coefficient and fiber location for SERT-Cre mice ($n = 10$). The red line is a linear regression curve, with its correlation coefficient shown as well ($p > 0.05$).
- (d) Correlation between Cox regression photostimulation coefficient and number of YFP expressing cells for SERT-Cre mice ($n = 10$). The red line is a linear regression curve, with its correlation coefficient shown as well ($p < 0.05$).

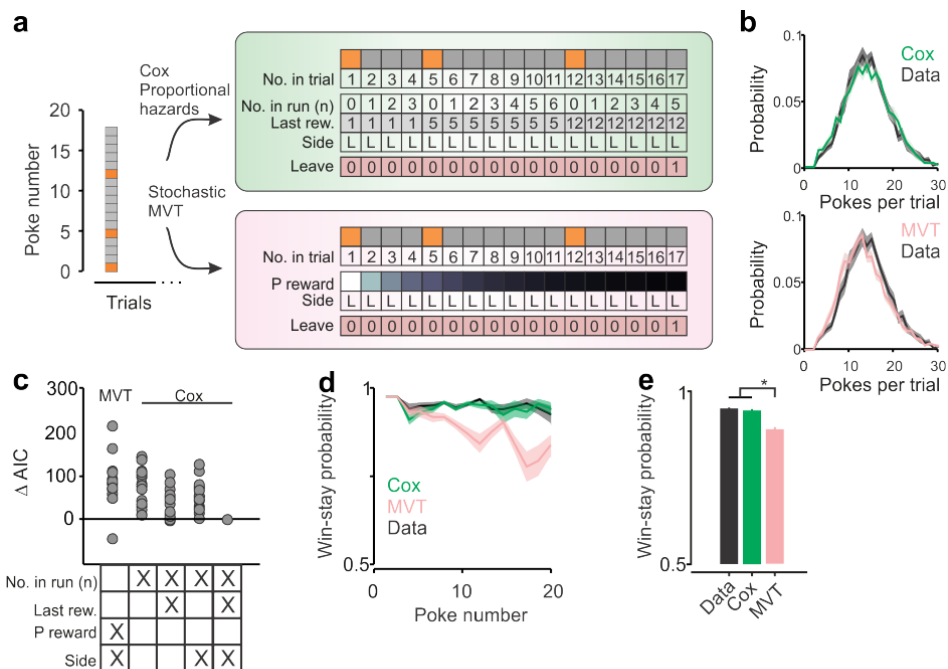


Supplementary figure 3 The effect of DRN 5-HT photostimulation on individual mouse switching behavior

(a-p) Distributions of the number of pokes per trial for photostimulated trials (blue for SERT-Cre mice and green for wild-type) and control trials (black) for all the mice used in this study.

(q) Histogram of Cox regression photostimulation coefficient for SERT-Cre (blue; $n = 10$) and wild-type (green; $n = 6$) mice. Filled bars correspond to significant coefficients ($p < 0.05$).

(r) Correlation between Cox regression photostimulation coefficient and average number of pokes per trial for SERT-Cre mice ($n = 10$). The red line is a linear regression curve, with its correlation coefficient shown as well ($p > 0.05$).



Supplementary figure 4 Model comparison

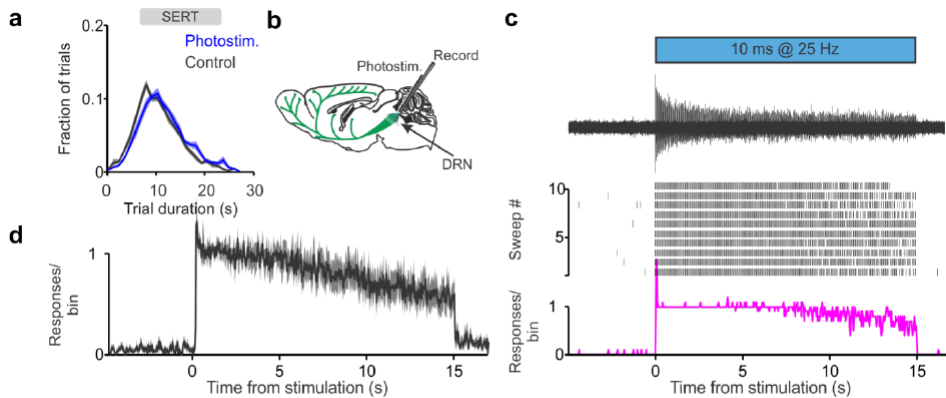
(a) Schematic drawing of model-fitting pipe-line for either the proportional hazards or stochastic MVT model. Top: in the proportional hazards model, each nose-poke was labeled with a vector of values corresponding to the different Cox regression coefficients. These values, together with the outcome of each poke (stay or leave) were used to fit a logistic regression model - the outcome of which was an estimated hazard rate that is reset at trial start and after each reward, and is multiplicatively changed by the different coefficient values. Bottom: in the stochastic MVT model, the subjective reward probability (calculate using Eq. 2 and depicted here with grayscale coloring) and side were used to fit a logistic regression model for mouse leaving decisions.

(b) Average pokes-per trial distributions for real and simulated data using the proportional hazards model (top; green) and the stochastic MVT model (bottom; pink).

(c) Akaike information criteria (AIC) for the stochastic MVT and four versions of the proportional hazards model. Regressors used in each of the models are indicated in the matrix below ($n = 16$ mice).

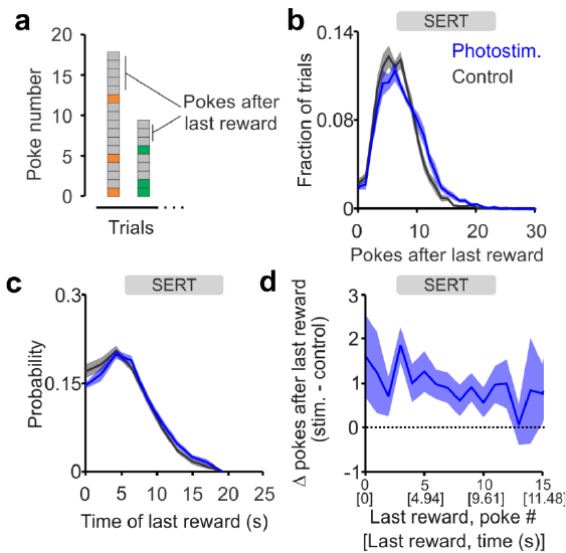
(d) Win-stay probability as a function of poke number within a trial for real and simulated data ($n = 16$ mice).

(e) Bar plot showing the average win-stay probability for the two models and data ($n = 16$ mice). $P < 0.05$, ANOVA.



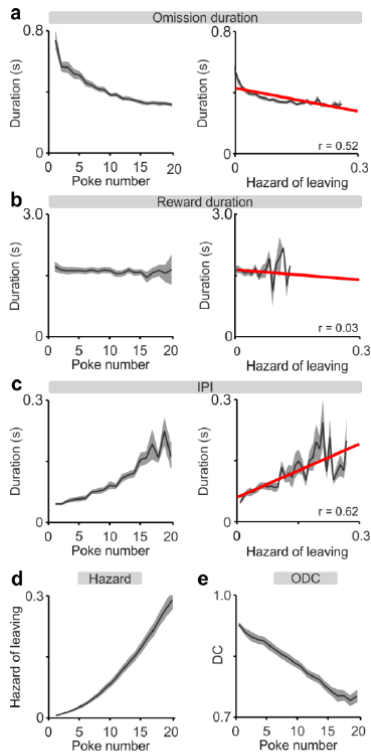
Supplementary figure 5 Electrophysiological validation of DRN 5-HT photostimulation

- (a) Distributions of trial durations for photostimulated (blue) and control trials (black) averaged across the population of SERT-Cre mice ($n = 10$).
- (b) Schematic of the experimental setup. Multi-unit responses to light delivery of ChR2-expressing DRN neurons were recorded using an optrode (adapted with permission from Ref. 63).
- (c) Top: Example multi-unit response to 15 s, 25 Hz photostimulation. Bottom: Mean responses of the same unit to 10 stimulation sweeps. Responses were binned in 40 ms bins aligned on light-pulse onset.
- (d) Average binned multi-unit responses to 15 s, 25 Hz photostimulation ($n = 7$ multi-units from 2 mice).



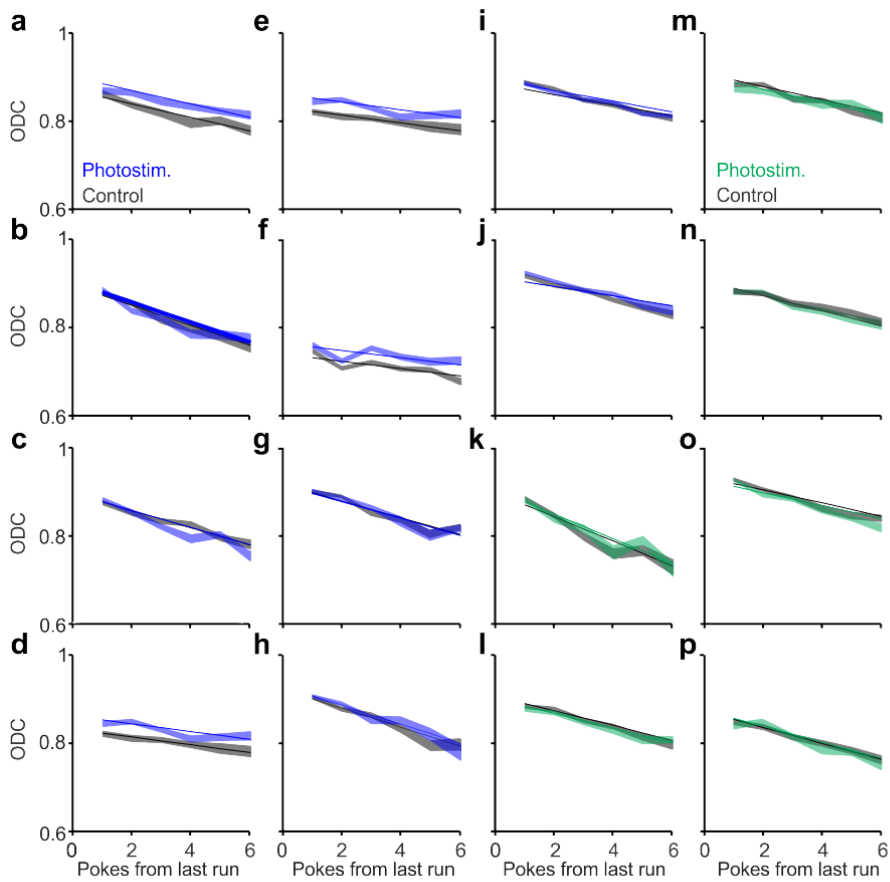
Supplementary Figure 6 The effect of DRN 5-HT photostimulation as a function of its duration

- (a) Two consecutive example trials, the first had five pokes after the last reward and the second had three.
- (b) Distributions of the number of pokes after last reward for photostimulated trials (blue) and control trials (black) averaged across the population of SERT-Cre mice ($p < 0.05$, Wilcoxon sign-rank test, $n = 10$).
- (c) Distributions of last reward times for photostimulated (blue) and control (black) trials across the population of SERT-Cre mice ($p < 0.05$, Wilcoxon sign-rank test, $n = 10$).
- (d) Difference between average number of pokes after last reward in photostimulated and control trials as a function of last reward position or time for SERT-Cre mice ($n = 10$). Last reward times are shown in brackets and are the average times of the rewards occurring at the corresponding position. $p > 0.05$, linear regression analysis.



Supplementary figure 7 Correlation between hazard and nose-poking behavior

- (a) Left: omission nose-poke duration as a function of nose-poke number within a trial. Right: correlation between omission duration and estimated hazard. The red line is a linear regression curve, with its equation and correlation coefficient shown as well ($p < 0.001$).
- (b) Left: Rewarded nose-poke duration as a function of nose-poke number within a trial. Right: correlation between reward duration and estimated hazard. The red line is a linear regression curve, with its equation and correlation coefficient shown as well ($p > 0.05$).
- (c) Left: Inter-poke-intervals duration as a function of nose-poke number within a trial. Right: correlation between inter-poke-interval duration and estimated hazard. The red line is a linear regression curve, with its equation and correlation coefficient shown as well ($p < 0.001$).
- (d) Estimated hazard as a function of nose-poke number within a trial.
- (e) ODC as a function of nose-poke number within a trial.



Supplementary figure 8 The effect of DRN 5-HT photostimulation on the microstructure of individual mouse behavior

(a-p) The ODC aligned on last reward for photostimulated trials (blue for SERT-Cre mice and green for wild-types) and control trials (black) for all the mice used in this study. The lines are linear regression curves.

Chapter 3: Serotonin elicits paradoxical effects on behavioral persistence and flexibility in foraging decisions

Dario Sarra, Eran Lottem & Zachary F. Mainen

Introduction

Serotonin (5-hydroxytryptamine, 5HT) has long been implicated in reward-guided behavior, but its precise role remains unclear. Behavioral flexibility is the capacity to shift action when it becomes ineffective due to changes in the environment, whereas behavioral persistence refers to the ability to maintain a previously rewarded action in the face of spurious adverse outcomes. These two aspects of reward-guided behavior have been traditionally studied separately. However, both are affected by serotonergic manipulations.

In reversal tasks, after pavlovian or operant conditioning, the stimulus- or action-outcome contingencies are changed covertly from the subjects. Repetitions of the previously rewarded action are labeled as perseverative errors, and changes in this quantity are used to measure behavioral flexibility. In these tasks, serotonin depletion (Clarke et al., 2004, 2005, 2007; Lapiz-Bluhm et al., 2009) or inhibition (Matias et al., 2017) results in significantly more perseverative errors, while pharmacological enhancement of 5-HT using SSRI (selective serotonin reuptake inhibitor, Brigman et al., 2010) reduces perseverative errors, suggesting that 5-HT promotes behavioral flexibility. However, evidence for contrasting roles of 5-HT_{2A} and 5-HT_{2C} receptors complicates the picture. Specifically, in rats performing a spatial reversal-learning task, systemic 5-HT_{2A} receptor antagonism impairs behavioral flexibility, while 5-HT_{2C} receptor antagonism facilitates it (Boulougouris et al., 2008). Therefore, although it is well established that 5-HT plays a role in behavioral flexibility, it is unclear why global 5-HT manipulation results in effects similar to those evoked by selectively acting on the 5-HT_{2A} receptor and in direct opposition to the effects of the 5-HT_{2C} receptor.

A separate line of studies focusing on behavioral persistence has suggested a different role for 5-HT, in apparent contradiction to the conclusion that 5-HT promotes behavioral flexibility. An essential feature of the tasks employed to study persistence is that the rules of the environment do not change after the stimuli- or action-outcome contingencies are learned. We will therefore refer to them collectively as “steady tasks.” In such contexts, where rewards depend on the ability to persist, in withholding or repeating an action, global serotonergic depletion (Eagle et al., 2009; Harrison et al., 1997a, 1997b, 1999) or inhibition (Miyazaki et al., 2012; Xu et al., 2017) significantly reduces animals’ persistence. Furthermore, optogenetic stimulation (Miyazaki et al., 2014; Fonseca et al., 2015; Xu et al., 2017; Ohmura et al., 2020) in these tasks increases animals’ persistence. These results appear in stark contradiction to those in the reversal tasks, where 5-HT activation reduced perseveration, and 5-HT inhibition increased it. Interestingly, just like in the reversal tasks, the 5-HT_{2A} and 5-HT_{2C} receptors have opposing effects in steady tasks, decreasing and promoting persistence, respectively (Winstanley et al., 2004; Fletcher et al., 2007). Overall, it appears that 5-HT promotes flexibility in reversal tasks, possibly through the 5HT_{2a} receptor, while it enhances persistence in steady tasks, perhaps through the 5-HT_{2C}.

Crucially, in the natural world, behavioral flexibility and persistence likely coexist in a continuum, and the balance between them determines behavior. Therefore we reasoned that a better understanding of the relationship between serotonin and flexibility/persistence would require a setting where both of these tendencies must be balanced to maximize reward. Contrary to reversal and steady tasks, which require either flexibility or persistence, foraging tasks represent an ecologically relevant approach whose optimal solution lies in the balance between

these two tendencies. In these paradigms, animals have to optimize for how long to exploit a depleting patch or bear a cost to explore an alternative, potentially more rewarding option. On the one hand, in this context, flexibility is necessary not to waste energy searching in an exhausted patch even if it proved to be rewarding in the recent past. On the other hand, persistence is required to avoid leaving a patch after spurious unsuccessful attempts, wasting energy to travel while the current patch is still sufficiently rewarding. The optimal tradeoff between these tendencies is formalized within foraging theory by the marginal value theorem (MVT, Charnov, 1976; Wajnberg et al., 2004; Lottem et al., 2018), which states that optimal foragers should leave a depleting resource whenever the instantaneous reward rate within a patch drops below the long-term average reward rate calculated taking into account travel times. Hence, we reckon that assessing how 5-HT manipulations affect the balance of persistence and flexibility within the MVT framework would lead to a better understanding of its role.

This study tested mice in a previously described two-ports probabilistic foraging task (Lottem et al., 2018). First, we found that systemic administration of 5-HT_{2A} or 5-HT_{2C} receptors antagonists reduces flexibility or persistence, respectively. This result is coherent with what was observed previously in reversal and steady tasks, suggesting that the 5-HT impact on these tasks and foraging involves overlapping mechanisms, at least partially. Second, we investigated how 5-HT could potentially drive either flexibility or persistence. In other ligand-receptor systems, such as noradrenaline, receptors that drive opposing effects are preferentially activated by different ligand concentrations (Shine et al., 2021) or exposure times (Lacey et al., 1993). We hypothesized that the temporal profile of 5-HT release could elicit different behavioral effects. We compared the effect of

optogenetically driven 5-HT transient release versus SSRI driven sustained increase in 5-HT levels to test this hypothesis. We found that transient activation promotes persistence in a foraging task, while sustained release promotes flexibility.

These results support the hypothesis that 5-HT's role in foraging is to balance persistence and flexibility. Furthermore, our study supports the hypothesis of a potential temporal mechanism that acts on different receptors. Future research should focus on understanding if the opposing effects could be caused by not timescale-related differences in the techniques, which is paramount to our understanding of how basic research translates to therapeutics.

Results

Animals engaging in a foraging task adopt a heuristic in line with the marginal value theorem solution

To study how 5-HT adjusts the balance between flexibility and persistence, we employed a probabilistic foraging task that we previously developed, described in Chapter 2 (Lottem et al., 2018). The behavioral box consisted of a rectangular chamber (50 cm long) with two water ports equipped with an infrared detector located on the short edges. Water-deprived mice foraged for water by probing the water port with their snout (Fig. 1a). Two foraging patches were defined as regions of interest (ROI) around each port through online video processing (Lopes et al., 2015, see Chapter 2 Methods). A foraging “trial” was defined as the period between entry into and exit from a foraging patch (Fig. 1b). Trials were considered correct if the animal alternates visits between the two patches, while consecutive visits to the same ROI were considered error trials and were not rewarded. Mice sought water reward probing

the ports via nose-poking, and exclusively in correct trials, each attempt was probabilistically rewarded or not with a water drop. The reward probability decayed exponentially with each consecutive nose-poke in the same port and was reset to an initial value at the start of a new correct trial (Fig. 1c). To verify that the animals foraged to balance the trade-off between exploration and exploitation coherently with the MVT predictions, we introduced three uncued randomly interleaved trial types with different initial probabilities of reward (easy-, medium-, and hard-quality patches; corresponding to 100%, 75% and 50% initial reward chance). Reward probability decayed exponentially at the same rate regardless of initial probability (Fig. 1d; Eq. 1 in Methods). Since the probability of reward decreases with each attempt, the amount of time necessary to obtain another reward increases to the point at which it is more valuable to spend time traveling to a fresh site than to persist longer.

As in Chapter 2 (Lottem et al., 2018) we tested if mice (29 C57BL/6 mice, 20 wild-types and 9 SERT-Cre) decide to leave a foraging patch in accordance with the MVT best policy (Charnov, 1976; Wajnberg et al., 2006). In agreement with our previous finding we observed that (1) animals leave the current patch as the reward rate is equal to the environment average reward rate (0.21 ± 0.008 reward s⁻¹ on average, and 0.20 ± 0.008 rewards s⁻¹ at leaving (mean $\pm$ SEM); $p = 0.535$, paired t-test, $n = 29$ mice, Fig. 2a), (2) at leaving time, the reward probability is the same independently of the patch quality (likelihood ratio test on $\text{RewardProbability} \sim 1 + \text{TrialType} + (1 \mid \text{MouseID})$ v.s. $\text{RewardProbability} \sim 1 + (1 \mid \text{MouseID})$; $p = 0.8$, Fig. 2b), (3) animals obtain the highest and lowest reward from the high and low-quality patches, respectively (likelihood ratio test on $\text{RewardNumber} \sim 1 + \text{TrialType} + (1 \mid \text{MouseID})$ v.s. $\text{RewardNumber} \sim 1 + (1 \mid \text{MouseID})$:

$p < 1e-99$, Fig. 2c), (4) animals make more attempts in high-quality patches and less in low-quality patches (likelihood ratio test on Pokes $\sim 1 + \text{TrialType} + (1 \mid \text{MouseID})$ v.s. Pokes $\sim 1 + (1 \mid \text{MouseID})$: $p < 1e-99$, Fig. 2d).

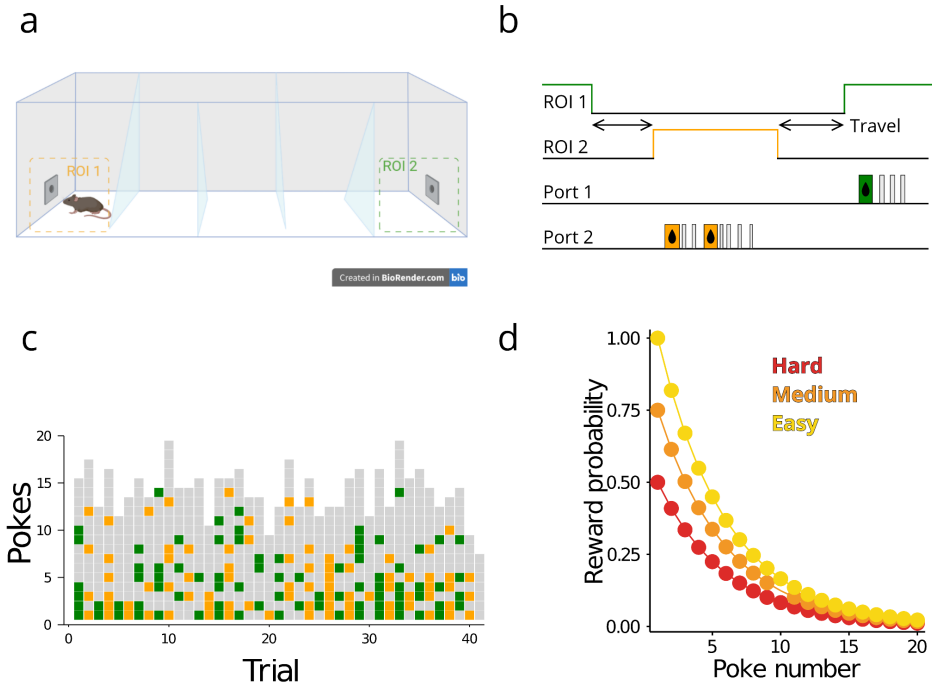


Figure 1. Animals forage adopting a heuristic in line with the MVT solution. (a) Schematic drawing of the foraging task apparatus. Water-deprived mice sought water reward alternating nose-poking in two reward ports at opposite short ends of a rectangular box. (b) Task events during behavior. A trial begins and ends upon entering and exiting a foraging site ROI, respectively. After reaching a foraging site, mice nose-poke multiple times and receive probabilistic rewards from the port before switching back to the opposite side. (c) Example session showing each trial's events in columns. Colored squares indicate rewarded attempts (orange for the left and green for the right port), and grey squares indicate unrewarded attempts. (d) Decreasing exponential functions from which reward probability is calculated for each consecutive attempt in a trial. Each function has a different scaling factor but the same time constant. The three functions define an Easy, Medium, and Hard trial-type depending on the magnitude of the initial reward probability (exponential scaling factor).

In our previous work (Chapter 2, Lottem et al., 2018) we showed that a proportional hazard model is a possible heuristic to perform the task in a way that is qualitatively similar to the MVT solution without actual knowledge of the reward probability. This approach models a leaving decision variable that resets at every reward and grows as a function of (1) number of attempts performed, (2) consecutive unrewarded attempts, and (3) last reward position in a trial. Measuring the relationship between the number of pokes after the last reward and the position of the last reward in each trial (Fig. 2e), we can qualitatively observe the dependence of the leaving decision on properties 2 and 3. As implied by the proportional hazard model we found that the amount of attempts after the last reward decays as a function of the last reward position (likelihood ratio test on $\text{AfterLastRewardPokes} \sim 1 + \text{LastRewardPosition} + (1 \mid \text{MouseID})$ v.s. $\text{AfterLastRewardPokes} \sim 1 + (1 \mid \text{MouseID})$: $p < 1e-3$, Fig. 2e). Consequently, if the animal were to adopt the proportional hazard model solution, it should perform more pokes after the last reward in the hard trial-type compared to the easy one, since the last reward would occur probabilistically sooner in hard and later in easy trials. We tested this hypothesis in our data and found that indeed the animals' behavior aligns with this qualitative prediction of the proportional hazard model (likelihood ratio test on $\text{AfterLastRewardPokes} \sim 1 + \text{TrialType} + (1 \mid \text{MouseID})$ v.s. $\text{AfterLastRewardPokes} \sim 1 + (1 \mid \text{MouseID})$: $p < 1e-14$, Fig. 2f).

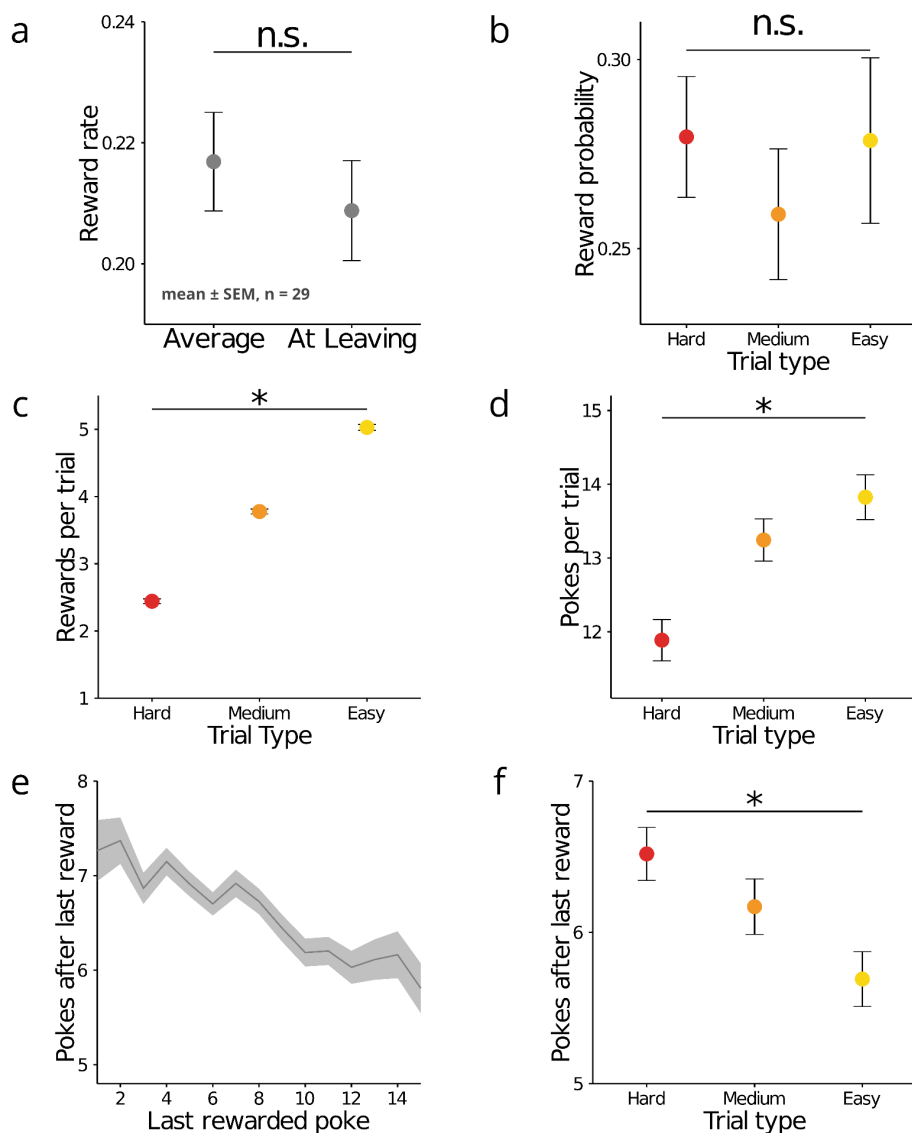


Figure 2. Animals forage by adopting a heuristic in line with the MVT solution. (a) Average reward rates throughout a session and only at the time of leaving a foraging patch. In agreement with the MVT, mice leave the patch when these two quantities match. (b) Mean reward probability across trial types at the time of leaving a foraging patch. As predicted by the MVT, this quantity is equivalent among the three trial types. (c) Average number of rewards obtained per trial type. As expected by MVT, mice obtain more water in easier trials. (d) Mean number of pokes (attempts) performed

before leaving depending on the trial-type. As per MVT, mice poke more times in easier than harder trial types. (e) Effects of last reward position on the number of pokes performed after the last reward before leaving a foraging patch. Although a reward resets the probability of leaving, the amount of further attempts is modulated by the total amount of attempts already performed. (f) Mean number of pokes after the last reward as a function of the trial type. As expected from the previous analysis, since the lower the probability of reward, the sooner the last reward in a trial is, mice perform more pokes after the last reward in the harder trials. This measure characterizes the relationship between the proportional hazard strategy and trial types.

In summary, we showed that taking advantage of the different trial-type design we can test for agreement with the previously described behavior testing for three properties: (1) the probability of reward at leaving is independent of the trial type, (2) the number of pokes per trial is positively correlated with the trial type initial probability of reward, and (3) the number of pokes after the last reward is negatively correlated with the trial type initial probability of reward.

Foraging task engages the 5-HT receptome in a similar manner to reversal and steady tasks

We hypothesized that the effect of 5-HT in foraging shares similar receptor substrates to reversal and steady tasks. We tested this by assessing the impact of 5-HT_{2A} (Altanserin, 0.5 mg/kg, i.p.) and 5-HT_{2C} (SB-242084, 3 mg/kg, s.c.) receptors' antagonism during foraging. Twenty wild-type C57BL/6 mice were trained in the foraging task and afterwards tested under pharmacological manipulation. Mice were divided into groups A and B, to deliver drugs on different weekdays to avoid confounds with potential weekly oscillations in thirst (see methods). To have a deeper understanding of the effect of 5-HT receptors on foraging, in our analysis, we aim to distinguish between qualitative and quantitative changes in the behavior. Qualitative changes indicate adopting a distinct behavioral strategy, while changes underlying

the same strategy would only present quantitative deviations. Since we have a well-characterized description of the effects of the trial type on three quantities, we can use linear mixed models to separately test for quantitative differences as main effects on the model, and qualitative changes, as interaction effects between the drugs and the trial-types.

First, we verified that animals behave in agreement with the previously described strategy. Indeed, after either 5-HT2a or 5-HT2c blockade, we observed no significant interaction effect between the treatment and the three qualitative properties of the behavior: (1) the probability of reward at leaving was equal for the three trial types (likelihood ratio test on $\text{RewardProbability} \sim 1 + \text{TrialType} + \text{Treatment} + (1 | \text{MouseID})$ v.s. $\text{RewardProbability} \sim 1 + \text{TrialType} + \text{Treatment} + \text{TrialType} \& \text{Treatment} + (1 | \text{MouseID})$: $p = 0.59$, Fig. 3a), (2) the number of attempts was positively correlated with the initial reward probability of the trial type (likelihood ratio test on $\text{Pokes} \sim 1 + \text{TrialType} + \text{Treatment} + (1 | \text{MouseID})$ v.s. $\text{Pokes} \sim 1 + \text{TrialType} + \text{Treatment} + \text{TrialType} \& \text{Treatment} + (1 | \text{MouseID})$: $p = 0.88$, Fig. 3b), and (3) the number of attempts after the last reward was negatively correlated with the initial reward probability of the trial type (likelihood ratio test on $\text{AfterLastRewardPokes} \sim 1 + \text{TrialType} + \text{Treatment} + (1 | \text{MouseID})$ v.s. $\text{AfterLastRewardPokes} \sim 1 + \text{TrialType} + \text{Treatment} + \text{TrialType} \& \text{Treatment} + (1 | \text{MouseID})$: $p = 0.74$, Fig. 3c).

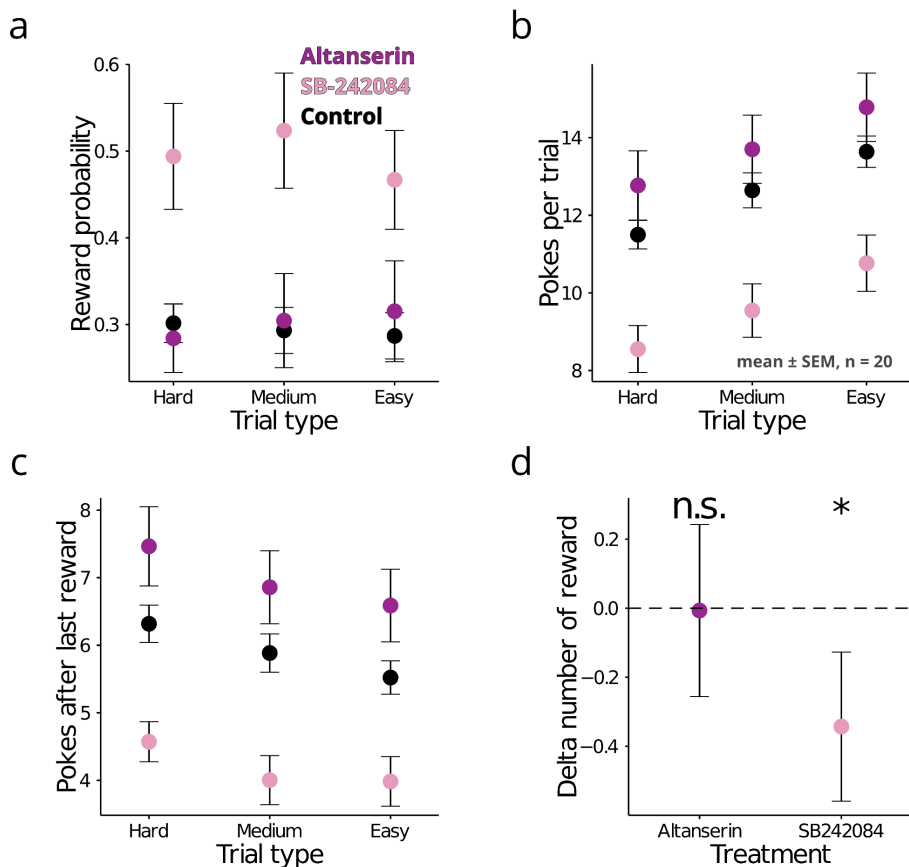


Figure 3. Effects of 5-HT2a and 5-HT2c antagonisms on foraging match the findings in reversal and steady tasks. (a) Average reward probability at the time of leaving per trial type after selective blockade of either 5-HT2a (Altanserin systemic injection, 0.5 mg/kg) or 5-HT2c receptors (SB-242084 systemic injection, 3mg/kg). The probability of reward at leaving is equal across trial types within all manipulation conditions, agreeing with the MVT prediction. (b) Mean number of pokes per trial. In accordance with previous results in reversal and steady tasks, 5-HT2a blockade reduces flexibility (significantly more pokes before leaving), and 5-HT2c blockade reduces persistence (significantly less pokes before leaving). However, these effects do not affect the MVT prediction that the higher the probability of reward, the higher the attempts before leaving. (c) Average number of pokes after the last reward per trial type. Similar to the previous analyses, 5-HT2a and 5-HT2c antagonists reduce flexibility and persistence, respectively, without affecting MVT predictions. (d) Difference in the average number of rewards obtained per trial (treatment - control). Altanserin increases the number of attempts performed without producing any significant increase

in rewards gained. SB-242084 reduces the pokes per trial but leads to a significant loss in rewards gained.

Next, we looked at quantitative changes produced by 5-HT receptor antagonism, with the specific intent to compare them with results from reversal and steady tasks. We found that blocking the 5-HT_{2A} and 5-HT_{2C} receptors, respectively increased and decreased the overall number of pokes (Pokes ~ 1 + Treatment + Protocol + (1 | MouseID), main effect of Altanserin: $p < 1e-7$, main effect of SB-242084: $p < 1e-38$, likelihood ratio test against Pokes ~ 1 + Protocol + (1 | MouseID): $p < 1e-57$, Fig. 3b) and the number of pokes after the last reward (AfterLastRewardPokes ~ 1 + Treatment + Protocol + (1 | MouseID), main effect of Altanserin: $p < 1e-12$, main effect of SB-242084: $p < 1e-28$, likelihood ratio test against AfterLastRewardPokes ~ 1 + Protocol + (1 | MouseID): $p < 1e-54$, Fig. 3c) performed before changing to the alternative patch. These results are in line with previous studies. As in the reversal tasks, we found that 5-HT_{2A} promotes flexibility, and its blockade increases the number of attempts before exploring an alternative option. Furthermore, as in steady tasks, we found that 5-HT_{2C} promotes persistence, and its blockade reduces the attempts endured before giving up.

Since using two sets of receptors, serotonin could skew behavior toward either persistence or flexibility; we postulated that the regular serotonergic activity should be optimized towards the best outcome. We assessed how performance was affected by 5-HT_{2a} and 5-HT_{2c} antagonism to test this hypothesis. On the one hand, although 5-HT_{2a} blockade made the mice stay longer, this did not increase the overall reward earned (paired t-test: $p = 0.95$, Fig. 3d), thus leading to a waste of energy compared to the normal behavior. On the other hand, blockade of 5-HT_{2c} receptors made the animal leave sooner, which led to a

significant loss in reward gained per trial (paired t-test: $p < 1e-4$, Fig. 3d). Thus, it appears that the endogenous activity of 5-HT_{2a} and 5-HT_{2c} receptors is balanced to maximize gains, and alterations in this balance are counterproductive in both directions.

In summary, our results suggest that the apparent paradoxical findings of serotonin promoting either flexibility or persistence in different contexts can be resolved by postulating a single role for 5-HT in balancing these two tendencies to produce adaptive behavior in the current context. In this perspective, the requirements of different tasks can be seen as being distributed along a continuum between persistence and flexibility. Thus, the paradoxical outcomes of global serotonergic manipulation can be interpreted as an effect of the task requirements that, in extreme cases (i.e., steady or reversal tasks), make only one of these two behavioral tendencies beneficial. However, one necessary prediction of this hypothesis is that 5-HT release should be able to promote both flexibility and persistence in the same task.

Sustained and transient serotonin produce paradoxical effects on foraging decisions

Our results on 5-HT_{2a} and 5-HT_{2c} manipulation suggest that foraging, reversal, and steady tasks share at least partially an equivalent serotonergic receptors substrate and that the system's activity is regulated to maximize gains. We thus sought to investigate how 5-HT release could control both tendencies within the same task/context to test this hypothesis. Previous electrophysiological recordings showed that the serotonergic neuron can transmit information with both tonic and phasic changes in their firing rate (Cohen et al., 2015; Hayashi et al.,

2015; Cohen and Grossman, 2020) and that their activity signals an unsigned prediction error (Matias et al., 2017). Interestingly, a key distinction in reversal and steady tasks regards the timescale of the uncertainty relevant to the studied decision. In steady tasks, some discrepancy between prediction and outcomes is expected and typically occurs within a trial with a timescale of seconds. In these cases, a tendency towards persistence helps maximize gains. Persistence prevents abandoning a strategy that provides stable outcomes, albeit probabilistic, due to meaningless fluctuations. By contrast, in reversal tasks, it is necessary to detect alterations in the environment states that require a change in strategy. These changes produce deviations from the expected uncertainty (unexpected uncertainty) throughout multiple trials on a typical time scale of tens of seconds or minutes. Thus, when these environment changes occur, a tendency towards flexibility results in exploring alternative solutions in the new state.

The centrality of the uncertainty duration is supported by previous studies on the potential computational mechanisms involved in distinguishing expected and unexpected uncertainty. Practically some mechanisms of evidence accumulation over trials are integral to unexpected uncertainty but not to the expected uncertainty (Doya, 2002; Dayan and Yu, 2002; Yu and Dayan, 2005). In the Reinforcement Learning framework, the ability to adjust decision making distinguishing between expected and unexpected uncertainty can be modeled as changes in the learning rate, “meta-learning,” and there is evidence that humans and other animals use this strategy (Behrens et al., 2007; Herzfeld et al., 2014; Massi et al., 2018; Soltani and Izquierdo, 2019).

Crucially serotonergic activity is known to adjust the learning rate with effects dependent on the neural activity (Grossman et al., 2020) or the behavior timescale (Igaya et al., 2018).

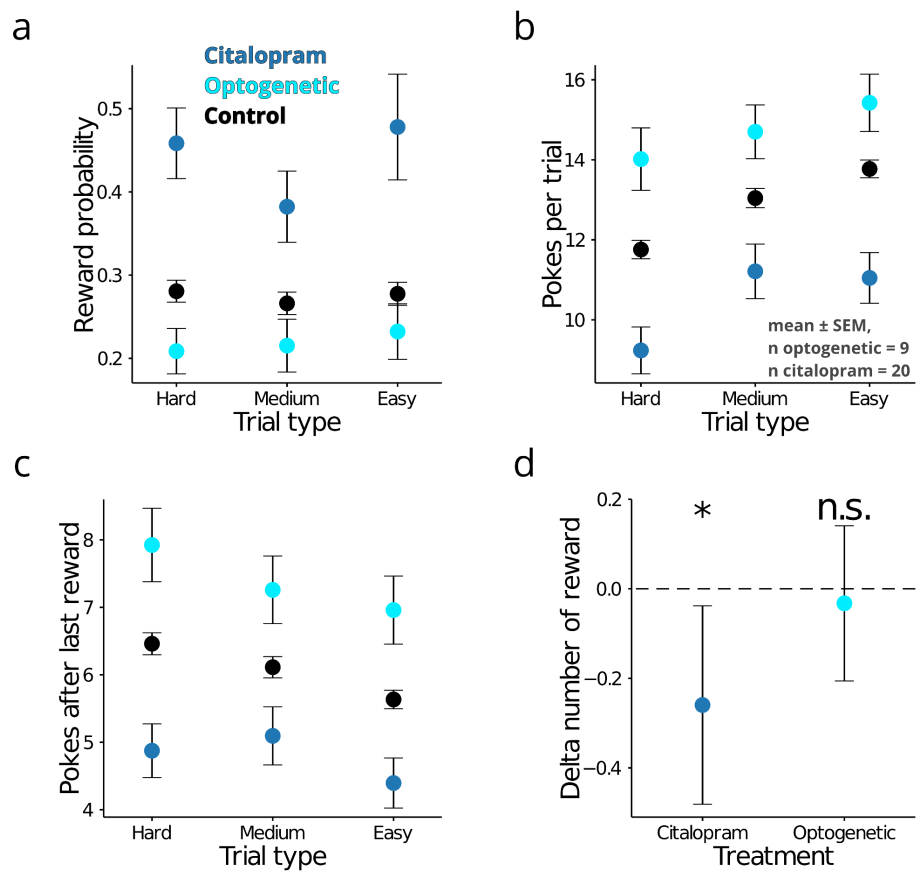


Figure 4. Effects of sustained and transient 5-HT release on foraging promote flexibility and persistence, respectively. (a) Average reward probability at the time of leaving per trial type after either sustained (Citalopram systemic injection, 5 mg/kg) or transient (5-HT DRN neurons optogenetic stimulation, 25Hz, 50% of trials) 5-HT release. Within all manipulation conditions, the probability of reward at leaving is equal across trial types, in agreement with the MVT prediction. (b) Mean number of pokes per trial. In accordance with the balancing hypothesis, sustained 5-HT release increases flexibility (significantly less pokes before leaving), while transient 5-HT release enhances persistence (significantly more pokes before leaving). However, these effects do not affect the MVT prediction that the higher the probability of reward,

the higher the attempts before leaving. (c) Average number of pokes after the last reward per trial type. Similar to the previous analyses, sustained and transient 5-HT release enhances flexibility and persistence, respectively, without affecting MVT predictions. (d) Difference in the average number of rewards obtained per trial (treatment - control). Result indicates that deviation from the learned balance between persistence and flexibility are detrimental in both directions. Citalopram systemic injection reduces the pokes per trial but leads to a significant loss in rewards gained. DRN 5-HT neurons' optogenetic stimulation increases the number of attempts performed without producing any significant increase in rewards gained.

We hypothesized that the timescale of 5-HT release could be a mechanism by which downstream neurons receptive to 5-HT adjust the balance between flexibility and persistence, with a sustained release favoring one tendency and a transient one favoring the other. Similar mechanisms where a ligand can exert preferential activation of one of its receptors over another, depending on concentration levels or exposure duration, have been well characterized in other pharmacological systems such as the Adrenaline and Noradrenaline systems (Lacey et al., 1993; Shine et al., 2021). Regarding the Adrenaline system, we know that the preferential activation of its opposing receptor is achieved thanks to affinity differences. Typically the receptor with lower affinity is preferentially activated by higher concentrations or more extended time exposures. In line with this feature, the 5-HT_{2a} and 5-HT_{2c} receptors have different affinities for 5-HT, with 5-HT_{2a} having lower affinity than 5-HT_{2c} (5-HT_{2a}: 6.0-6.5 percentage inhibition, binding affinity (pKi); Egan et al., 1998; Rothman et al., 2000; Shapiro et al., 2000; 5-HT_{2c}: 6.7-7.9 pKi; Herrick-Davis et al., 1997; Rothman et al., 2000). Thus, if 5-HT were to preferentially activate 5-HT_{2a} or 5-HT_{2c} receptor depending on its release duration, we would expect that a sustained 5-HT release has opposite effects of 5-HT_{2a} antagonist, while a transient 5-HT release has the opposite effects of 5-HT_{2c} antagonists. To test this hypothesis, we compared the effects of two 5-HT release modes during foraging, mimicking transient and sustained serotonergic

activation using optogenetic stimulation (ChR2) and serotonin reuptake inhibitor (SSRI, Citalopram), respectively. Although both of these techniques have been shown to elicit 5-HT release measured with in vivo microdialysis in the frontal cortex (David et al., 2003; Miyazaki et al., 2014), the systemic injection of SSRIs produces a 5-HT increase in the scale of one hour, while it is in the scale of seconds for the optogenetic stimulation of dorsal raphe 5-HT neurons.

We tested the effect of SSRI administration (Citalopram, 5mg/kg, i.p.) during foraging on 20 C57BL/6 wild-type animals using the same protocol described above. Optogenetic stimulation was performed on a separate cohort of 9 SERT-Cre mice as described in Chapter 2 (Lottem et al., 2018). As for 5-HT_{2a} and 5-HT_{2c} antagonism we found no qualitative change in the strategy adopted by the animal after 5-HT release with either optogenetic activation or SSRI. Linear mixed models analysis revealed no significant interaction between trial type and treatment in terms of the three key features of the behavior: (1) the probability of reward at leaving was equal for the three trial types (likelihood ratio test on RewardProbability $\sim 1 + \text{TrialType} + \text{Treatment} + (1 | \text{MouseID})$ v.s. RewardProbability $\sim 1 + \text{TrialType} + \text{Treatment} + \text{TrialType} \& \text{Treatment} + (1 | \text{MouseID})$: $p = 0.46$, Fig. 4a), (2) the number of attempts was positively correlated with the initial reward probability of the trial type (likelihood ratio test on Pokes $\sim 1 + \text{TrialType} + \text{Treatment} + (1 | \text{MouseID})$ v.s. Pokes $\sim 1 + \text{TrialType} + \text{Treatment} + \text{TrialType} \& \text{Treatment} + (1 | \text{MouseID})$: $p = 0.33$, Fig. 4b), and (3) the number of attempts after the last reward was negatively correlated with the initial reward probability of the trial type (likelihood ratio test on AfterLastRewardPokes $\sim 1 + \text{TrialType} + \text{Treatment} + (1 | \text{MouseID})$ v.s. AfterLastRewardPokes $\sim 1 + \text{TrialType} + \text{Treatment} + \text{TrialType} \& \text{Treatment} + (1 | \text{MouseID})$: $p = 0.52$, Fig. 4c).

Next we verified that both the optogenetic and SSRI mediated 5-HT release led to significant quantitative changes in behavior. On the one hand, as shown previously (Lottem et al., 2018), transient 5-HT release through optogenetic stimulation led to an increase in the overall and after last reward number of pokes before leaving (Fig. 4b). On the other hand, a sustained release of 5-HT, through Citalopram injections, led to a significant reduction of both measures (Pokes $\sim 1 + \text{Treatment} + \text{Protocol} + (1 | \text{MouseID})$, main effect of Optogenetic: $p < 1e-12$, main effect of Citalopram: $p < 1e-14$, likelihood ratio test against Pokes $\sim 1 + \text{Protocol} + (1 | \text{MouseID})$: $p < 1e-24$; AfterLastRewardPokes $\sim 1 + \text{Treatment} + \text{Protocol} + (1 | \text{MouseID})$, main effect of Optogenetic: $p < 1e-17$, main effect of Citalopram: $p < 1e-9$, likelihood ratio test against AfterLastRewardPokes $\sim 1 + \text{Protocol} + (1 | \text{MouseID})$: $p < 1e-24$, Fig. 4c).

Finally, we tested if the normal serotonergic levels were optimized to maximize gains as assessed above. We found that deviation from the spontaneous 5-HT release was detrimental in both cases, with optogenetic activation leading to more attempts without any further gains (paired t-test: $p = 0.67$, Fig. 4d), and citalopram reducing the number of attempts at the cost of rewards obtained per trial (paired t-test: $p < 0.05$, Fig. 4d).

Our results support the hypothesis that 5-HT can lead to both flexibility or persistence within the same context and that sustained and transient release in the downstream circuitry may control which tendency predominates at any given moment. Furthermore, we show that in normal conditions, 5-HT levels appear to be balanced to maximize reward gains from the environment.

Discussion

This study aimed to offer a potential interpretation to reconcile the 5-HT paradoxical effects on decision-making tasks: specifically, the contradictory findings indicating that 5-HT promotes either flexibility - the capacity to shift action when it becomes ineffective due to changes in the environment - or persistence - the ability to maintain a previously rewarded action in the face of spurious adverse outcomes. We hypothesized that 5-HT's role is to balance these two tendencies rather than promote one of them selectively and that part of the contradiction arises from task designs where only one of them is beneficial. To this aim, we tested the effects of 5-HT_{2a} and 5-HT_{2c} antagonisms in a foraging setting where the best strategy requires a balance between persistence and flexibility. First, our findings confirm that 5-HT_{2a} promotes flexibility, while 5-HT_{2c} enhances persistence, in line with previous results from both reversal and steady tasks (Winstanley et al., 2004; Fletcher et al., 2007; Boulougouris et al., 2008). This result confirms that this subset of the 5-HT's receptome has consistent effects in a spectrum of task demands with flexibility and persistence at the extremes. Crucially, we proved that changes towards any of these tendencies impoverish the acquired behavioral strategy, validating the hypothesis that their natural activity was adjusted to balance flexibility and persistence and maximize gains.

We next investigated how flexibility and persistence could be alternatively favored by 5-HT release and tested the hypothesis that the balance between these two downstream effects could be controlled by the 5-HT release concentration or duration, similarly to other ligand-receptor systems (Lacey et al., 1993; Shine et al., 2021). We found that indeed a sustained release of 5-HT, produced by SSRI,

promotes flexibility, while a transient release, elicited by optogenetic stimulation, enhances persistence. In agreement with our results from manipulating the 5-HT_{2a} and 5-HT_{2c} receptors, both optogenetic and citalopram deviate the behavior from a well-adjusted balance, resulting in either an increase in expenses or a loss in gains.

How may these results explain the differences seen in reversal and steady tasks? One interesting difference between these classes of tasks is the temporal dynamic of their challenges. Although both contexts (i) involve reward-guided behaviors, (ii) manipulate action-outcome contingencies, and (iii) use perseverance as a measure of interest, they produce different forms of variability. Computationally, they manipulate two distinct sources of uncertainty: expected uncertainty, the known unreliability of action-outcome links within a familiar environment, and unexpected uncertainty, resulting from substantial violations of expectations underlying qualitative changes in the environment (Doya, 2002; Dayan and Yu, 2002; Yu and Dayan, 2005; Gross et al., 2020). Importantly, they do so while operating on different timescales. On the one hand, in reversal tasks, the action- or stimulus-outcome rules change to produce an unexpected uncertainty event only after the subject has learned the contingencies. Thus, the detection of changes in the environment occurs at relatively slow time scales across blocks of trials that span tens of seconds or minutes. On the other hand, in steady tasks, since the environment rules are fixed, the variability is only manipulated in terms of delay, magnitude or probability of the reward, resulting in expected uncertainty oscillations across trials that span only seconds. Previous studies have shown that 5HT neurons track variability in the environment, suggesting that these neurons carry signals related to unsigned prediction errors (Matias et al., 2017) or uncertainty (Grossman et al., 2020). Crucially, recordings in the

Dorsal Raphe Nucleus from either unidentified (Hayashi et al., 2015) or serotonergic neurons (Cohen et al., 2015) have shown that their tonic and phasic firing independently encode behaviorally relevant variables. A possible interpretation is that, on the one hand, prediction errors in steady tasks occur transiently, and to overcome spurious negative outcomes a transient serotonin release promotes persistence, potentially through the 5HT_{2C} receptors. On the other hand, in the face of prolonged prediction errors, such as in reversal tasks, a prolonged increase in serotonin would promote flexibility, possibly through its 5HT_{2A} receptors.

Although our results are necessary requirements to support this hypothesis, they are not sufficient. An essential advantage of using selective antagonists in interrogating the circuitry involved in decision making is that the cause of their effects is primarily restricted to removing the contribution of the ordinarily occurring 5-HT's release. Thus we can generally exclude that this approach affects circuits that are not typically recruited in the process. However, both 5-HT_{2a} and 5-HT_{2c} are constitutively active, and antagonists such as Altanserin and SB-242084 can have diverse effects on the different signal transduction pathways (Barnes and Sharp, 1999; Herrick-Davis, 1999; Berg et al., 2001-2005). To better clarify how 5-HT modulates flexibility and persistence through these two receptors, future studies should investigate the relationship between the different transduction pathways and their behavioral effects.

Furthermore, 5-HT_{2a} and 5-HT-2c expressions in the brain have some unique spatial distributions (Mengod et al., 2010), although partially overlapping. Therefore, it is uncertain if they exert their opposing effects acting in the same downstream areas. Interrogating local circuits through local pharmacological injection has led to a complex picture with mixed results depending on the behavioral context. For instance, local

injection of 5-HT_{2c} antagonist in the orbitofrontal cortex (OFC), but not the medial prefrontal cortex (mPFC) or the nucleus accumbens (NAc), reduces persistence in a spatial reversal task (Boulougouris and Robbins, 2010). By contrast, local 5-HT_{2a} antagonist infusion did not affect a spatial reversal task in any of those regions (Boulougouris and Robbins, 2010). However, after local injection of 5-HT_{2a} antagonists in OFC, reversal deficits have been reported in an attention-shifting task (Furr et al., 2012). Regarding steady tasks, in the 5-choice serial reaction time task (5CSRTT), 5-HT_{2a} and 5-HT_{2c} local injection in the NAc, but not the prelimbic or infralimbic cortex, have opposing effects (Robinson et al., 2008) in line with what was observed in our task. In order to support the hypothesis of a balancing effect through 5-HT_{2a} and 5-HT_{2c} opponence, future research will have to investigate the effects of their antagonism on foraging with spatially selective manipulations.

The sustained and transient hypothesis explaining the opposing effects of 5-HT release due to optogenetic or SSRI manipulation will also need to be verified with local manipulation studies. Systemic SSRI injection affects the whole brain, while our optogenetic stimulation is restricted to areas targeted by serotonergic neurons in the Dorsal Raphe Nucleus. Although the Venn diagram of the targeted area will have significant overlap, we cannot exclude that the opposing effects could be caused by activation of different downstream areas. Furthermore, the optogenetic stimulation of sets of neurons within the DRN projecting to different targets (Ren et al., 2018) can elicit opposing effects on behavior. It is also essential to consider that the two techniques present different off-target effects. While the SSRI Citalopram has a highly selective effect (Hyttel, 1982), optogenetic stimulation produces the corelease of other ligands besides 5-HT in particular glutamate, and it is known that some of the effects of the 5-HT DRN neurons' optogenetic

stimulation are actually due to glutamate release (Sengupta et al., 2017; Wang et al., 2019). Altogether, these considerations underscore that our current results support an opposing role for sustained and transient 5-HT release; further studies focused on specific 5-HT downstream targets will be necessary to exclude other potential explanations.

Notwithstanding all of the above, this study presents a new functional perspective on the role of 5-HT that encompasses behavioral results on decision-making typically separated. Such a reunifying framework offers a promising testable benchmark to identify the role of serotonin and potentially its relevance in various neuropsychiatric disorders. Interestingly in the action selection field, it has been proposed that action control emerges from setting the balance between persistence and flexibility (Hommel, 2015). This is particularly relevant in light of the evidence suggesting that 5-HT activity is linked to the perceived controllability over the environment (Amat 1998; Maswood 1998; Grahn et al., 1999; Amat et al., 2005; Robbins 2005). With recent findings revealing context-dependent activity (Seo, Guru, et al., 2019) and quantitative links between 5-HT and adaptation to the environment changes (Grossman, 2020), this work supports the emerging concept that the 5-HT role is to balance behavioral response to produce adaptive behavior.

Materials and Methods

Animal subject: Twenty 3-4 months old C57BL/6 mice were used in systemic pharmacology experiments: Altanserin, SB-242084 (Fig. 3), and Citalopram (Fig. 4). To reduce animal usage, we employed wild-type animals from different genetic backgrounds (8 Dat-Cre; 5 Gad2-Cre; 5 Sert-Cre; 1 VGAT-ChR2; 1 F512-Cre). Ten heterozygous Sert-Cre mice were employed for optogenetic manipulation of DRN serotonergic

neurons (Fig. 4). One animal was excluded from the study due to the loss of its implant. The Dat-Cre mouse line was obtained from the Jackson Laboratory (stock number: 006660). The Gad2-Cre was obtained from the Jackson Laboratory (stock number: 010802). The Sert-Cre mouse line 61 was obtained from the Mutant Mouse Regional Resource Centers (stock number: 017260-UCD). The VGAT-ChR2 mouse line 8 was obtained from the Jackson Laboratory (stock number: 014548). The Fl12-Cre mouse line was obtained from the Mutant Mouse Regional Resource Centers (stock number: 017262-UCD). All experimental procedures were approved and performed in accordance with the Champalimaud Centre for the Unknown Ethics Committee guidelines and by the Portuguese Veterinary General Board (Direcao-Geral de Veterinaria, approval 0421/000/000/2016). The mice were kept under a normal 12 h light/dark cycle, and training, as well as testing, occurred during the light period. Before testing or after surgeries, for the inactivation experiments, mice were single-housed. During training and testing, the mice were water-deprived; mice performed 1 session per day, 5 days a week. Water was available to them only during task performance in training or testing days. In the 2 days of testing break, they received first 800ul, then 600ul of water. Food was freely accessible to the mice in their home cages. Extra water was provided if needed to ensure that mice maintain no less than 80% of their original weight.

Adeno-associated virus injection and cannula implantation: Mice were anesthetized with isoflurane (4% induction and 0.5–1% for maintenance) and placed in a stereotaxic frame (David Kopf Instruments, Tujunga, CA). Lidocaine (2%) was injected subcutaneously before incising the scalp. In order to infect serotonergic neurons with ChR2, a craniotomy was drilled over the cerebellum and a pipette filled

with AAV2.9.EF1a.DIO.hChR2(H134R)-eYFP.WPRE.hGH 1013GCmL-1 viral solution (University of Pennsylvania) was lowered to the DRN (Bregma -4.7 AP, -2.9 DV) with a 32° angle toward the back of the animal. The viral solution (1 µL) was injected using a Picospritzer II (Parker). After waiting for 10–15 min, the pipette was removed from the brain, and an optical fiber (200 µm core diameter, 0.48 NA, 4–5 mm long, Doric lenses) was lowered through the same craniotomy such that its tip was positioned 200 µm above the injection point. The implant was cemented to the skull using dental acrylic (Pi-Ku-Plast HP 36, Bredent, Senden, Germany). Mice were monitored until recovery from the surgery and returned to their home cages. Gentamicin (48760, Sigma-Aldrich, St. Louis, MO) was topically applied around the implant.

Drugs and Injections: All drugs were purchased from Tocris (Bio-Techne Ltd., Abingdon, United Kingdom). Animals were divided into four groups receiving treatment in different weekdays counterbalanced in two behavioral boxes, as shown in the table below. Control data (marked in blue font) used in all the analyses were taken from the behavioral session before the treatment (marked in red) when the specific drug vehicle was injected.

group	box	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
A	1	none	VEH	DRUG	VEH	SAL	homecage	homecage
B	2	none	SAL	VEH	DRUG	VEH	homecage	homecage
C	1	none	SAL	VEH	DRUG	VEH	homecage	homecage
D	2	none	VEH	DRUG	VEH	SAL	homecage	homecage

Drug vehicle, concentration, dosage, route of administration, and time of injection before the behavioral testing are summarized in the table below. Note that to avoid delivering large amounts of fluid injections that could disrupt the motivation to forage, we made solutions with high

drug concentrations that would lead to an average injection for a mouse of 25 grams equal to 0.15 ml.

Drug	Vehicle	Drug concentration	Dose	Route	Time before testing
Altanserin	Sal + 1% DMSO	0.1 mg/ml	0.5 mg/kg	I.P.	40 min
SB242084	Sal + 3% DMSO	0.6 mg/ml	3 mg/kg	S.C.	40 min
Citalopram	Sal	1 mg/ml	5 mg/kg	I.P.	20 min

Optogenetic stimulation: In order to optically stimulate ChR2-expressing 5-HT neurons, we used blue light from a 473 nm laser (LRS-0473-PFF-00800-03, Laserglow Technologies, Toronto, Canada or DHOM-M-473-200, UltraLasers, Inc., Newmarket, Canada) that was controlled by an acousto-optical modulator (AOM; MTS110-A1-VIS or MTS110-A3- VIS, AA optoelectronic, Orsay, France). Light exiting the AOM was focused into an optical fiber patchcord (200 μ m, 0.22 NA, Doric lenses), connected to a second fiber patchcord through a rotary joint (FRJ 1 $\times$ 1, Doric lenses), which was then connected to the chronically implanted optic fiber cannula. The photostimulation protocol consisted of a train of 10 ms wide pulses at 25 Hz, measuring 5 mW at fiber tip, triggered by the first nose-poke in a trial and lasted until the trial's end (i.e., by either ROI exit, or if 10 s had elapsed since the last nose-poke exit).

Data and statistical analysis: Behavioral data analysis was performed using custom-written scripts in Julia-1.4.1. Behavioral results in figures were represented as mean $\pm$ standard error of the mean (SEM) across mice, and statistical significance was accepted for p-values < 0.05. The statistical analysis was done in Julia-1.4.1 with the HypothesisTests and MixedModels existing packages. Paired group differences (mean

number of rewards per trial) were tested using a paired t-test. The effect of trial type, manipulation, and interaction between them was tested by applying a linear mixed-effects model regression. To validate the relevance of the experimental manipulations (trial type or manipulation) and their interaction, we compared nested models (a general model and a special case model, excluding or including the factor of interest, respectively) using a likelihood ratio test. The test corresponds to a chi-squared test on the difference of the deviance of the two nested models, with degrees of freedom equal to the difference in degrees of freedom between the general model (lacking the predictor) and its special case (with the predictor of interest). For each analysis, we report the test statistics. We use Wilkinson annotation to describe the models with denoting random effects.

References

- Amat, J., Baratta, M. V., Paul, E., Bland, S. T., Watkins, L. R., & Maier, S. F. (2005). Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nature Neuroscience*, 8(3), 365–371. <https://doi.org/10.1038/nn1399>
- Amat, J., Matus-Amat, P., Watkins, L. R., & Maier, S. F. (1998). Escapable and inescapable stress differentially alter extracellular levels of 5-HT in the basolateral amygdala of the rat. *Brain Research*, 812(1–2), 113–120. [https://doi.org/10.1016/S0006-8993\(98\)00960-3](https://doi.org/10.1016/S0006-8993(98)00960-3)
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9), 1214–1221. <https://doi.org/10.1038/NN1954>
- Berg, K. A., Stout, B. D., Maayani, S., & Clarke, W. P. (2001). Differences in Rapid Desensitization of 5-Hydroxytryptamine 2A. *Pharmacology*, 299(2), 593–602.
- Berg, K. A., Harvey, J. A., Spampinato, U., & Clarke, W. P. (2005). Physiological relevance of constitutive activity of 5-HT_{2A} and 5-HT_{2C} receptors. *Trends in Pharmacological Sciences*, 26(12), 625–630. <https://doi.org/10.1016/j.tips.2005.10.008>

Boulougouris, V., Glennon, J. C., & Robbins, T. W. (2008). Dissociable effects of selective 5-HT_{2A} and 5-HT_{2C} receptor antagonists on serial spatial reversal learning in rats. *Neuropsychopharmacology*, 33(8), 2007–2019. <https://doi.org/10.1038/sj.npp.1301584>

Boulougouris, V., & Robbins, T. W. (2010). Enhancement of Spatial Reversal Learning by 5-HT_{2C} Receptor Antagonism Is Neuroanatomically Specific. *Journal of Neuroscience*, 30(3), 930–938. <https://doi.org/10.1523/JNEUROSCI.4312-09.2010>

Brigman, J. L., Mathur, P., Harvey-White, J., Izquierdo, A., Saksida, L. M., Bussey, T. J., ... Holmes, A. (2010). Pharmacological or genetic inactivation of the serotonin transporter improves reversal learning in mice. *Cerebral Cortex*, 20(8), 1955–1963. <https://doi.org/10.1093/cercor/bhp266>

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, 9(2), 129–136. [https://doi.org/10.1016/0040-5809\(76\)90040-X](https://doi.org/10.1016/0040-5809(76)90040-X)

Clarke, H. F., Dalley, J. W., Crofts, H. S., Robbins, T. W., & Roberts, A. C. (2004). Cognitive Inflexibility after Prefrontal Serotonin Depletion. *Science*, 304(5672), 878–880. <https://doi.org/10.1126/science.1094987>

Clarke, H. F., Walker, S. C., Crofts, H. S., Dalley, J. W., Robbins, T. W., & Roberts, A. C. (2005). Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. *Journal of Neuroscience*, 25(2), 532–538. <https://doi.org/10.1523/JNEUROSCI.3690-04.2005>

Clarke, H. F., Walker, S. C., Dalley, J. W., Robbins, T. W., & Roberts, A. C. (2007). Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. *Cerebral Cortex*, 17(1), 18–27. <https://doi.org/10.1093/cercor/bhj120>

Cohen, J. Y., Amoroso, M. W., & Uchida, N. (2015). Serotonergic neurons signal reward and punishment on multiple timescales. *eLife*, 2015(4), 1–25. <https://doi.org/10.7554/eLife.06346>

Cohen, J. Y., & Grossman, C. D. (2020). Dorsal raphe serotonergic neurons regulate behavior on multiple timescales. *Handbook of Behavioral Neuroscience* (Vol. 31). Elsevier. <https://doi.org/10.1016/B978-0-444-64125-0.00030-X>

David, D. J. P., Bourin, M., Jegu, G., Przybylski, C., Jolliet, P., & Gardier, A. M. (2003). Effects of acute treatment with paroxetine, citalopram and venlafaxine in vivo on noradrenaline and serotonin outflow: A microdialysis study in Swiss mice. *British Journal of Pharmacology*, 140(6), 1128–1136. <https://doi.org/10.1038/sj.bjp.0705538>

Dayan, P., & Yu, A. J. (2002). Expected and Unexpected Uncertainty: ACh and NE in the Neocortex. *Advances in Neural Information Processing Systems*, 15.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, 15(4–6), 495–506. [https://doi.org/10.1016/S0893-6080\(02\)00044-8](https://doi.org/10.1016/S0893-6080(02)00044-8)

Eagle, D. M., Lehmann, O., Theobald, D. E. H., Pena, Y., Zakaria, R., Ghosh, R., ... Robbins, T. W. (2009). Serotonin depletion impairs waiting but not stop-signal reaction time in rats: Implications for theories of the role of 5-HT in behavioral inhibition. *Neuropsychopharmacology*, 34(5), 1311–1321. <https://doi.org/10.1038/npp.2008.202>

Egan, C. T., Herrick-Davis, K., & Teitler, M. (1998). Creation of a Constitutively Activated State of the 5-Hydroxytryptamine_{2A} Receptor by Site-Directed Mutagenesis: Inverse Agonist Activity of Antipsychotic Drugs. *Journal of Pharmacology and Experimental Therapeutics*, 286(1).

Fletcher, P. J., Tampakeras, M., Sinyard, J., & Higgins, G. A. (2007). Opposing effects of 5-HT_{2A} and 5-HT_{2C} receptor antagonists in the rat and mouse on premature responding in the five-choice serial reaction time test. *Psychopharmacology*, 195(2), 223–234. <https://doi.org/10.1007/s00213-007-0891-z>

Fonseca, M. S., Murakami, M., & Mainen, Z. F. (2015). Activation of dorsal raphe serotonergic neurons promotes waiting but is not reinforcing. *Current Biology*, 25(3), 306–315. <https://doi.org/10.1016/j.cub.2014.12.002>

Furr, A., Lapiz-Bluhm, M. D., & Morilak, D. A. (2012). 5-HT_{2A}-receptors in the orbitofrontal cortex facilitate reversal learning and contribute to the beneficial cognitive effects of chronic citalopram treatment in rats. *The International Journal of Neuropsychopharmacology / Official Scientific Journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)*, 15(9), 1295. <https://doi.org/10.1017/S1461145711001441>

Grahn, R. E., Will, M. J., Hammack, S. E., Maswood, S., McQueen, M. B., Watkins, L. R., & Maier, S. F. (1999). Activation of serotonin-immunoreactive cells in the dorsal raphe nucleus in rats exposed to an uncontrollable stressor. *Brain Research*, 826(1), 35–43. [https://doi.org/10.1016/S0006-8993\(99\)01208-1](https://doi.org/10.1016/S0006-8993(99)01208-1)

Grossman, C. D., Bari, B. A., & Cohen, J. Y. (2020). Serotonin neurons modulate learning rate through uncertainty. *BioRxiv*, 2020.10.24.353508. <https://doi.org/10.1101/2020.10.24.353508>

Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1999). Central serotonin depletion impairs both the acquisition and performance of a symmetrically reinforced go/no-go conditional visual discrimination. *Behavioural Brain Research*, 100(1–2), 99–112. [https://doi.org/10.1016/S0166-4328\(98\)00117-X](https://doi.org/10.1016/S0166-4328(98)00117-X)

Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1997). Doubly dissociable effects of median- and dorsal-raphe lesions on the performance of the five-choice serial reaction time test of attention in rats. *Behavioural Brain Research*, 89(1–2), 135–149. [https://doi.org/10.1016/S0166-4328\(97\)00053-3](https://doi.org/10.1016/S0166-4328(97)00053-3)

- Harrison, A. A., Everitt, B. J., & Robbins, T. W. (1997). Central 5-HT depletion enhances impulsive responding without affecting the accuracy of attentional performance: Interactions with dopaminergic mechanisms. *Psychopharmacology*, 133(4), 329–342. <https://doi.org/10.1007/s002130050410>
- Hayashi, K., Nakao, K., & Nakamura, K. (2015). Appetitive and aversive information coding in the primate dorsal raphe nucleus. *Journal of Neuroscience*, 35(15), 6195–6208. <https://doi.org/10.1523/JNEUROSCI.2860-14.2015>
- Herrick-Davis, K., Egan, C., & Teitler, M. (1997). Activating mutations of the serotonin 5-HT_{2C} receptor. *Journal of Neurochemistry*, 69(3), 1138–1144. <https://doi.org/10.1046/J.1471-4159.1997.69031138.X>
- Herrick-Davis, K., Grinde, E., & Niswender, C. M. (1999). Serotonin 5-HT_{2C} Receptor RNA Editing Alters Receptor Basal Activity. *Journal of Neurochemistry*, 73(4), 1711–1717. <https://doi.org/10.1046/j.1471-4159.1999.731711.x>
- Herzfeld, D. J., Vaswani, P. A., Marko, M. K., & Shadmehr, R. (2014). A memory of errors in sensorimotor learning. *Science (New York, N.Y.)*, 345(6202), 1349–1353. <https://doi.org/10.1126/SCIENCE.1253138>
- Hommel, B. (2015). *Between Persistence and Flexibility*. BS:ADMS (Vol. 2). Elsevier. <https://doi.org/10.1016/bs.adms.2015.04.003>
- Hyttel, J. (1982). Citalopram — Pharmacological profile of a specific serotonin uptake inhibitor with antidepressant activity. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 6(3), 277–295. [https://doi.org/10.1016/S0278-5846\(82\)80179-6](https://doi.org/10.1016/S0278-5846(82)80179-6)
- Iigaya, K., Fonseca, M. S., Murakami, M., Mainen, Z. F., & Dayan, P. (2018). An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nature Communications*, 9(1). <https://doi.org/10.1038/S41467-018-04840-2>
- Lacey, R. J., Cable, H. C., James, R. F. L., London, N. J. M., Scarpello, J. H. B., & Morgan, N. G. (1993). Concentration-dependent effects of adrenaline on the profile of insulin secretion from isolated human islets of Langerhans. *Journal of Endocrinology*, 138(3), 555–563. <https://doi.org/10.1677/joe.0.1380555>
- Lapiz-Bluhm, M. D. S., Soto-Piña, A. E., Hensler, J. G., & Morilak, D. A. (2009). Chronic intermittent cold stress and serotonin depletion induce deficits of reversal learning in an attentional set-shifting test in rats. *Psychopharmacology*, 202(1–3), 329–341. <https://doi.org/10.1007/s00213-008-1224-6>
- Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., Lohuis, M. oude, & Mainen, Z. F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature Communications*, 9(1), 1000. <https://doi.org/10.1038/s41467-018-03438-y>

- Massi, B., Donahue, C. H., & Lee, D. (2018). Volatility Facilitates Value Updating in the Prefrontal Cortex. *Neuron*, 99(3), 598-608.e4. <https://doi.org/10.1016/J.NEURON.2018.06.033>
- Maswood, S., Barter, J. E., Watkins, L. R., & Maier, S. F. (1998). Exposure to inescapable but not escapable shock increases extracellular levels of 5-HT in the dorsal raphe nucleus of the rat. *Brain Research*, 783(1), 115–120. [https://doi.org/10.1016/S0006-8993\(97\)01313-9](https://doi.org/10.1016/S0006-8993(97)01313-9)
- Matias, S., Lottem, E., Dugué, G. P., & Mainen, Z. F. (2017). Activity patterns of serotonin neurons underlying cognitive flexibility. *ELife*, 6, 1–24. <https://doi.org/10.7554/eLife.20552>
- Mengod, G., Cortés, R., Vilaró, M. T., & Hoyer, D. (2010). Distribution of 5-HT Receptors in the Central Nervous System. In *Handbook of Behavioral Neuroscience* (Vol. 21, pp. 123–138). Elsevier. [https://doi.org/10.1016/S1569-7339\(10\)70074-6](https://doi.org/10.1016/S1569-7339(10)70074-6)
- Miyazaki, K. W., Miyazaki, K., & Doya, K. (2012). Activation of Dorsal Raphe Serotonin Neurons Is Necessary for Waiting for Delayed Rewards. *Journal of Neuroscience*, 32(31), 10451–10457. <https://doi.org/10.1523/JNEUROSCI.0915-12.2012>
- Miyazaki, K. K. W., Miyazaki, K. K. W., Tanaka, K. F., Yamanaka, A., Takahashi, A., Tabuchi, S., & Doya, K. (2014). Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. *Current Biology*, 24(17), 2033–2040. <https://doi.org/10.1016/j.cub.2014.07.041>
- Ohmura, Y., Tsutsui-Kimura, I., Sasamori, H., Nebuka, M., Nishitani, N., Tanaka, K. F., ... Yoshioka, M. (2020). Different roles of distinct serotonergic pathways in anxiety-like behavior, antidepressant-like, and anti-impulsive effects. *Neuropharmacology*, 167(July), 107703. <https://doi.org/10.1016/j.neuropharm.2019.107703>
- Ren, J., Friedmann, D., Xiong, J., Liu, C. D., Ferguson, B. R., Weerakkody, T., ... Luo, L. (2018). Anatomically Defined and Functionally Distinct Dorsal Raphe Serotonin Sub-systems. *Cell*, 175(2), 472-487.e20. <https://doi.org/10.1016/j.cell.2018.07.043>
- Robbins, T. W. (2005). Controlling stress: how the brain protects itself from depression. *Nature Neuroscience*, 8(3), 261–262. <https://doi.org/10.1038/nn0305-261>
- Robinson, E. S., Dalley, J. W., Theobald, D. E., Glennon, J. C., Pezze, M. A., Murphy, E. R., & Robbins, T. W. (2008). Opposing roles for 5-HT_{2A} and 5-HT_{2C} receptors in the nucleus accumbens on inhibitory response control in the 5-choice serial reaction time task. *Neuropsychopharmacology*, 33(10), 2398–2406. <https://doi.org/10.1038/sj.npp.1301636>

Rothman, R. B., Baumann, M. H., Savage, J. E., Rauser, L., McBride, A., Hufeisen, S. J., & Roth, B. L. (2000). Evidence for Possible Involvement of 5-HT_{2B} Receptors in the Cardiac Valvulopathy Associated With Fenfluramine and Other Serotonergic Medications. *Circulation*, 102(23), 2836–2841. <https://doi.org/10.1161/01.CIR.102.23.2836>

Sengupta, A., Bocchio, M., Bannerman, D. M., Sharp, T., & Capogna, M. (2017). Control of Amygdala Circuits by 5-HT Neurons via 5-HT and Glutamate Cotransmission. *The Journal of Neuroscience*, 37(7), 1785–1796. <https://doi.org/10.1523/JNEUROSCI.2238-16.2016>

Seo, C., Guru, A., Jin, M., Ito, B., Sleezer, B. J., Ho, Y.-Y., ... Warden, M. R. (2019). Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode. *Science*, 542(February), 538–542. <https://doi.org/10.1126/science.aau8722>

Shapiro, D. A., Kristiansen, K., Kroeze, W. K., & Roth, B. L. (2000). Differential Modes of Agonist Binding to 5-Hydroxytryptamine_{2A} Serotonin Receptors Revealed by Mutation and Molecular Modeling of Conserved Residues in Transmembrane Region 5. *Molecular Pharmacology*, 58(5), 877–886. <https://doi.org/10.1124/MOL.58.5.877>

Shine, J. M., Müller, E. J., Munn, B., Cabral, J., Moran, R. J., & Breakspear, M. (2021). Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics. *Nature Neuroscience* 2021 24:6, 24(6), 765–776. <https://doi.org/10.1038/s41593-021-00824-6>

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews. Neuroscience*, 20(10), 635–644. <https://doi.org/10.1038/S41583-019-0180-Y>

Wajnberg, E., Bernhard, P., Hamelin, F., & Boivin, G. (2006). Optimal patch time allocation for time-limited foragers. *Behavioral Ecology and Sociobiology*, 60(1), 1–10. <https://doi.org/10.1007/s00265-005-0131-7>

Wang, H.-L., Zhang, S., Qi, J., Wang, H., Cachope, R., Mejias-Aponte, C. A., ... Morales, M. (2019). Dorsal Raphe Dual Serotonin-Glutamate Neurons Drive Reward by Establishing Excitatory Synapses on VTA Mesoaccumbens Dopamine Neurons. *Cell Reports*, 26(5), 1128–1142.e7. <https://doi.org/10.1016/J.CELREP.2019.01.014>

Winstanley, C. A., Theobald, D. E. H., Dalley, J. W., Glennon, J. C., & Robbins, T. W. (2004). 5-HT_{2A} and 5-HT_{2C} receptor antagonists have opposing effects on a measure of impulsivity: Interactions with global 5-HT depletion. *Psychopharmacology*, 176(3–4), 376–385. <https://doi.org/10.1007/s00213-004-1884-9>

Xu, S., Das, G., Hueske, E., & Tonegawa, S. (2017). Dorsal Raphe Serotonergic Neurons Control Intertemporal Choice under Trade-off. *Current Biology*, 27(20), 3111–3119.e3. <https://doi.org/10.1016/j.cub.2017.09.008>

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4), 681–692. <https://doi.org/10.1016/j.neuron.2005.04.026>

Author Contributions

D.S., E.L., and Z.F.M. designed the experiments and analyses. D.S. conducted the experiments. D.S. analyzed the data. D.S. wrote the original draft.

Chapter 4: Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas

Pietro Vertechì, Eran Lottem, Dario Sarra, Beatriz Godinho, Isaac Treves, Tiago Quendera, Matthijs Nicolai Oude Lohuis & Zachary F. Mainen

Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas

Pietro Vertechi,^{1,7} Eran Lottem,^{2,7} Dario Sarra,^{1,7} Beatriz Godinho,^{1,3} Isaac Treves,⁴ Tiago Quendera,¹ Matthijs Nicolai Oude Lohuis,^{5,6} and Zachary F. Mainen^{1,8,*}

¹Champalimaud Research, Champalimaud Centre for the Unknown, 1400-038 Lisbon, Portugal

²The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, 91904 Jerusalem, Israel

³Nuffield Department of Clinical Neurosciences, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

⁴MIT Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 46-2005, Cambridge, MA 02139-4307, USA

⁵Cognitive and Systems Neuroscience Group, Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam, 1098XH Amsterdam, the Netherlands

⁶Research Priority Area Brain and Cognition, University of Amsterdam, Amsterdam, the Netherlands

⁷These authors contributed equally

⁸Lead Contact

*Correspondence: zmainen@neuro.fchampalimaud.org

<https://doi.org/10.1016/j.neuron.2020.01.017>

SUMMARY

Essential features of the world are often hidden and must be inferred by constructing internal models based on indirect evidence. Here, to study the mechanisms of inference, we establish a foraging task that is naturalistic and easily learned yet can distinguish inference from simpler strategies such as the direct integration of sensory data. We show that both mice and humans learn a strategy consistent with optimal inference of a hidden state. However, humans acquire this strategy more than an order of magnitude faster than mice. Using optogenetics in mice, we show that orbitofrontal and anterior cingulate cortex inactivation impacts task performance, but only orbitofrontal inactivation reverts mice from an inference-based to a stimulus-bound decision strategy. These results establish a cross-species paradigm for studying the problem of inference-based decision making and begins to dissect the network of brain regions crucial for its performance.

INTRODUCTION

In natural foraging behaviors, animals must continually choose between trying to exploit resources at their current location and leaving to explore another, potentially superior one, at the expense of a possibly costly travel period. Viewed from the perspective of optimal decision-making, the crucial question is when is it best to leave the current site for another one? According to the marginal value theorem, in order to maximize returns, an optimal forager ought to leave its current site when

the immediate rate of reward drops below the average rate (Chamov, 1976). However, this elegant solution to the foraging problem only applies in deterministic environments (Kolling et al., 2014), in which both immediate and average reward rates are knowable to the agent. In a more realistic scenario—for example, where rewards are encountered probabilistically—the immediate reward rate is ill-defined and the marginal value theorem does not apply.

One widely used and powerful approach to model decision making in dynamic, stochastic environments is reinforcement learning (RL) (Sutton and Barto, 2018). In RL, the values of different actions (such as leaving a foraging site or staying on) are continuously updated through trial and error, based on their outcomes, allowing agents to adaptively modify their preferences as conditions change. In its simplest form, model-free RL assigns each action with a value that is updated based on its immediate outcome, with no regard to the causal, and often hidden, structure that links actions to outcomes. Although computationally efficient and consistent with a large body of experimental data on both Pavlovian and operant tasks (Eshel et al., 2016; Schultz et al., 1997), model-free RL is not the best available strategy in many situations. Consider, for instance, a lion that has just successfully captured prey. If the fact that in doing this it has most likely scared away all other animals is ignored, the lion may continue to hunt in the same region, wasting a considerable amount of time searching for the now long-gone prey. Conversely, things may turn out badly for a zebra if it assumes that its current foraging ground was safe (that is, lion-free) just because it had not seen a lion yet in its immediate surroundings. What these examples illustrate is that relying solely on recent outcomes, while ignoring causal structures in the world, may have suboptimal (if not catastrophic) consequences. Instead, in structure learning (Boyan et al., 1999; Braun et al., 2010; Pearl, 1991), a form of inference-based RL, agents choose actions based on their beliefs about the current state of the

world, which is determined by both incoming sensory evidence (such as outcomes) and knowledge of the underlying causal structure of the environment. How humans and animals implement such strategies remains an important and poorly understood question (Daw et al., 2011; Niv et al., 2015; Starkweather et al., 2018). The study of the neural mechanisms underlying flexible, integrative behavior has drawn special attention to the prefrontal cortex and the computational role of its different areas. Although the mapping of the rodent prefrontal cortex has not reached a consensus, here, we adopt the description of Uylings and van Eden (1990), which defines as rat prefrontal cortex those areas comparable to the primate prefrontal cortex in terms of thalamic reciprocal connections, corticocortical connections, and functional aspects, including the orbitofrontal cortex (OFC) and the anterior cingulate cortex (ACC). It has been suggested that the OFC is crucial for hidden state representation, and hence for inference-based decisions. For example, in both rats and primates, lesions or inhibition of OFC impairs subjects' ability to adjust their behavior in reversal learning tasks, where the depletion of a previously rewarding site (or the futility of a previously rewarding action) may be viewed as a change in the (hidden) state of the world (Wilson et al., 2014), even though this result may be technique-dependent: Rudebeck et al. (2013) found that aspiration lesions (that also damage passing fibers), but not excitotoxic lesions, of monkey OFC impaired reversal learning. The adjacent ACC, often considered part of the rodent medial prefrontal cortex (Starkweather et al., 2018; Tervo et al., 2014), has been implicated in monitoring value during foraging (Hayden et al., 2011; Kolling et al., 2012) and could be responsible for encoding the value of alternative options (Kolling et al., 2016) and changing behavior based on the decreasing value of the current option (Shima and Tanji, 1998; Williams et al., 2004).

Here, we describe a foraging task in which subjects may seek rewards at either one of two foraging sites. This task has a special hidden structure: at any given moment, only one of the sites can deliver rewards and the site of the rewards switches with a certain probability after each foraging attempt. Importantly, even when reward is available, it is not delivered for every attempt, but rather with a probability less than 1. This makes the task a partially observable Markov decision process (POMDP): the true state of world (i.e., the identity of the rewarding site) is hidden and subjects must infer it based on noisy observations. A defining feature of this task, due to the hidden structure, is the asymmetry of the evidence provided by rewards and failures (unrewarded attempts): a single failure provides partial evidence in favor of a site switch, whereas a single reward provides full certainty that the current site is rewarding. A "stimulus-bound" agent, in the sense of Wilson et al. (2014), would assign value to observable states (being on the left or on the right or the other foraging site) by linearly combining rewards and failures. Such a process does not capture the essential asymmetry of the task. Ten rewards are much better than one reward in terms of value, but under optimal inference, one single reward is as informative as ten, because it already gives absolute certainty that the current site is active. Thus, leaving decisions under stimulus-bound and inference-based strategies in this task will be qualitatively

different. A stimulus-bound agent will become more persistent the more rewards it has received at a site, whereas an inference-based agent will not show such an effect. We found that both mice and humans display hallmarks of inference in the performance of a foraging task and are able to build a non-trivial representation of task space. We further show that optogenetic inhibition of the OFC in mice selectively disrupts optimal inference behavior, biasing mice toward a sub-optimal stimulus-bound strategy. Similar inhibition of the adjacent ACC results in delayed leaving decisions but does not disrupt the inference process itself, suggesting a specific role of OFC in this important cognitive function.

RESULTS

A Probabilistic Foraging Task Can Dissociate Value or Evidence Accumulation

We developed a self-paced probabilistic foraging task. Subjects sought rewards by actively probing a foraging site. Each try at the active site yielded reward with probability p_{RWD} , and could cause a switch with probability p_{SW} (Figure 1A). After a state switch, to obtain more rewards, subjects needed to travel to a second site at some distance and therefore bear a travel cost. Subjects were thus tasked with inferring a hidden state of the current site through a sequence of observations of stochastic events (rewards and failures). There are actually many ways of integrating rewards and failures to form a decision. In a stimulus-bound process, the relative value of the left site with respect to the right site $V = V_{LEFT} - V_{RIGHT}$ would increase gradually with left rewards, decrease gradually with right rewards, and decay to 0 with failures (Figure 1B). In formulas, given a decay coefficient γ , a reward indicator r_t , a site indicator s_t (1 for left, -1 for right), and signed outcomes $o_t = r_t \times s_t$:

$$V_{t+1} = (1 - \gamma)V_t + \gamma o_{t+1}. \quad (\text{Equation 1})$$

On the other hand, an agent that is aware of the structure of the task—the fact that a hidden state determines which site is rewarding at any time—could use rewards and failures differently, allowing it to better infer whether the current foraging site is active or inactive. The relative value would then be:

$$V_t = p_{RWD}(P(\text{LeftActive}|r_1, s_1, \dots, r_t, s_t) - P(\text{RightActive}|r_1, s_1, \dots, r_t, s_t)), \quad (\text{Equation 2})$$

(see STAR Methods for a detailed treatment of the probability computation). Unlike the stimulus-bound mechanism in Figure 1B, this process is able to track effectively the rapidly evolving value of the foraging sites (Figure 1C). Both accumulation processes can be used as generative models of the behavior by defining the probability of staying on, e.g., the left site as a sigmoidal function of the relative value:

$$P(\text{NextLeft} | V, s) = \sigma(\beta(V + s \cdot T)), \quad (\text{Equation 3})$$

where β represents a softmax parameter (the higher it is, the more deterministic the behavior), and T represents the staying bias (when the value of the left and the right site are estimated

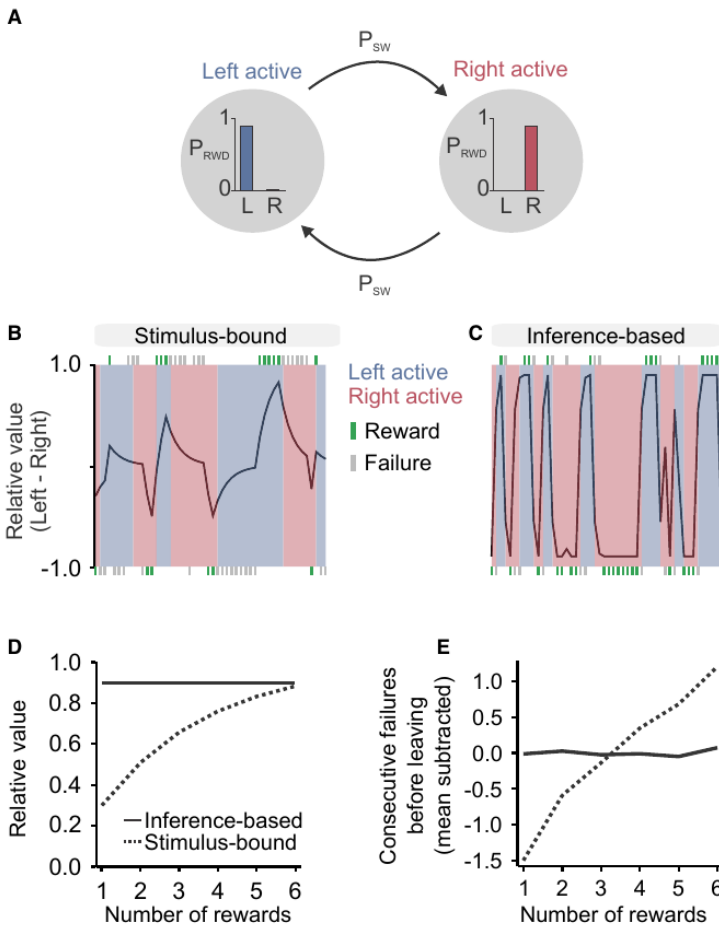


Figure 1. A Probabilistic Foraging Task Can Dissociate Stimulus-Bound from Inference-Based Evidence Accumulation

(A) Formally, the task is a hidden Markov model with *LeftActive* and *RightActive* states. It has two parameters: probability of reward given state and probability of state transition.

(B and C) Estimated relative value (left minus right) as a function of trial history (rewards in green, failures in gray) in the stimulus-bound model (B) and inference-based model (C), respectively. Shaded patches indicate actual state.

(D) Effect of rewards on relative value in stimulus-bound and inference-based models: the two models are simulated in a trial with only rewards on the same site. Relative value increases with reward number in the stimulus-bound but not in the inference-based model.

(E) Consecutive failures before leaving (normalized subtractively) as a function of reward number in a simulated data of stimulus-bound and inference-based models: reward number has an effect on consecutive failures in the stimulus-bound but not in the inference-based model.

12 days in a baseline protocol with $p_{RWD} = 0.9$ and $p_{SW} = 0.3$ and observed the effect of rewards on behavior during learning.

Because our task was a foraging style task, the mapping to choice and feedback are not present in a trial. Each trial contains a number of pokes, each of which contains feedback in the form of water (or not), as illustrated in Figure 2B. The choice of the mouse is whether to continue to poke (exploit) at the current site or whether to leave (explore), a choice it makes after each poke. The only incorrect choice is to leave to the

as equal, the subject should still prefer to stay to avoid the travel cost).

Both models predict that the probability of leaving increases with the number of consecutive failures. However, the effect of a reward is very different between them. In a stimulus-bound model, the probability of leaving decreases with the number of rewards, as each reward contributes to the accumulated value. In the inference-based model, it does not, because a single reward is sufficient to deduce with certainty that the current site is active (Figure 1D). Thus, a simple test of whether subjects are using inference is to check whether the number of failures before leaving changes with the number of preceding rewards (Figure 1E).

Mice Accumulate Evidence and Not Rewards

We first developed the hidden state foraging task as a rodent behavioral task (Figure 2A) in which mice had to nose-poke at one of two possible ports to obtain water rewards (2 μ L each). We trained 18 C57BL/6 wild-type mice of 2 months age for

other site before the state has switched. When the mouse switches site too early, no rewards will be emitted by the other port; the mouse is obliged to return to the original port and continue to poke. Mice made only $\sim 2.25\% \pm 0.47\%$ ($n = 18$ mice) errors on average. An example of the behavior of a trained animal is shown in Figure 2C (see Figure S1 for summary statistics of the durations of the various task epochs).

Mice tended to alternate bouts of pokes at a given site (6.98 ± 0.14 pokes per bout, $n = 18$ mice) with trips to the opposite site, producing a natural segmentation in trials (i.e., poke bouts on the same site). This presumably reflects the clear asymmetry in time cost between nose-poking again on the same site, a very cheap action (Figure S1, inter-poke interval = 0.16 ± 0.025 s, unrewarded poke duration = 0.33 ± 0.006 s, $n = 18$ mice), and switching site, a much more expensive option (Figure S1, 3.15 ± 0.18 s, $n = 18$ mice).

We found that the number of consecutive failures since the last reward (ConsecutiveFailureIndex) was a better predictor

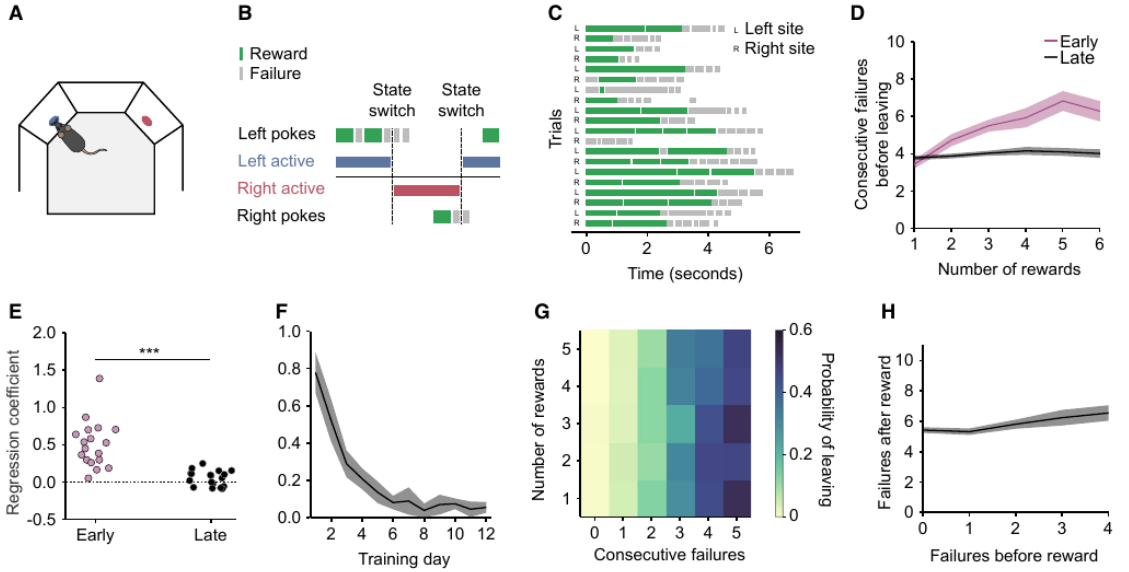


Figure 2. Mice Accumulate Inferred Evidence for State Switches and Not Site Value

(A) Schematic of rodent task. Mice shuttle back and forth between two reward sites to obtain water rewards.
 (B) Example sequence of pokes. Pokes in the correct site can be rewarded or not, whereas pokes in the incorrect site are never rewarded. Following a state switch, the animals need to travel to the other site to obtain more rewards.
 (C) Example behavior: sequence of poke bouts (i.e., trials) with rewards in green and failures in gray.
 (D) Consecutive failures before leaving as a function of reward number in early training (days 1 to 3, purple) compared with late training (days 10 to 12, black). Solid line represents mean and sha.
 (E) Slope coefficient in $ConsecutiveFailures \sim 1 + RewardNumber$ for early training and late training. Slope coefficient is higher in early trials, likelihood ratio test on linear mixed-effect model $ConsecutiveFailures \sim 1 + RewardNumber + Early + RewardNumber \& Early + (1|MouseID)$ versus a null model with no interaction: $p < 1e-10$, $n = 18$ mice (see [STAR Methods](#) for a description of the formula notation).
 (F) Evolution of reward number coefficient across days. Solid line and shaded area represent mean and across animals. Solid line and shaded area represent mean and SEM across animals.
 (G) Probability of leaving as a function of number of rewards and consecutive failures in late training.
 (H) Failures after reward as a function of failure before reward in trials with only one reward in a more difficult protocol. Solid line and shaded area represent mean and across animals. Solid line and shaded area represent mean and SEM across animals.
 See also [Video S1](#).

of mouse choice than the time spent at the nose poke ($TimeSpentAtPort$) ([Lottem et al., 2018](#)). We fitted two logistic regression models with random effects. Here, and throughout the text, we use Wilkinson notation ([Wilkinson and Rogers, 1973](#)) (see [STAR Methods](#) for a detailed explanation):

$$LeavingPort \sim 1 + ConsecutiveFailureIndex + (1|MouseID)$$

and

$$LeavingPort \sim 1 + TimeSpentAtPort + (1|MouseID).$$

The $ConsecutiveFailureIndex$ model was overwhelmingly better (deviance = 10,892) than the $TimeSpentAtPort$ model (deviance = 14,771). As confirmation, we also tested a model that included both predictors:

$$Leaving \sim 1 + ConsecutiveFailureIndex + TimeSpentAtPort + (1|MouseID),$$

and only the $ConsecutiveFailureIndex$ had a positive coefficient (0.78 ± 0.013) whereas the $TimeSpentAtPort$ had a small negative coefficient (-0.046 ± 0.006).

In the early part of training, animals were unaware of the structure of the task and exhibited hallmarks of a stimulus-bound strategy: more failures were needed to leave the foraging port in rich foraging bouts, with many rewards before a state switch, compared to poor foraging bouts, with as little as one reward before a state switch. After training, however, the number of rewards had no effect on the number of failures before leaving, consistent with an inference-based strategy ([Figure 2D](#)). To quantify this effect at a single animal level, we fitted a linear regression model that predicted the number of consecutive failures before leaving as a function of the number of prior rewards in the current trial (i.e., foraging bout at a given site): $ConsecutiveFailures \sim 1 + RewardNumber$ ([Figure 2E](#)). The data show that during the first days of training, there was a strong positive correlation between these two quantities, but

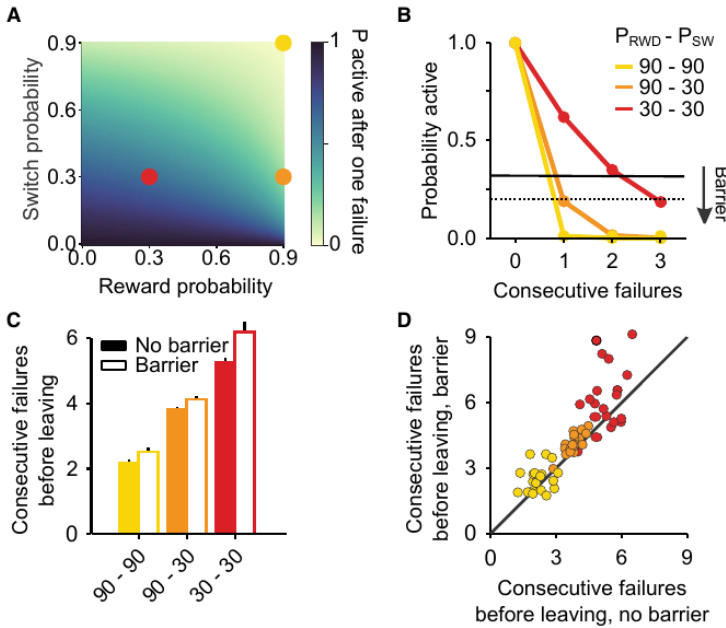


Figure 3. Accumulation of Inferred Evidence Is Tuned to Task Parameters

(A) Probability of being on the correct site after a failure as a function of reward probability and transition probability.

(B) Probability of being on the correct site as a function of trial history for three protocols (Easy environment: $p_{RWD} = 0.9$ and $p_{SW} = 0.9$; Medium environment: $p_{RWD} = 0.9$ and $p_{SW} = 0.3$; Hard environment: $p_{RWD} = 0.3$ and $p_{SW} = 0.3$). Leaving decisions can be modeled by setting a threshold on this probability that changes as a function of the travel cost (black lines).

(C) Consecutive failures before leaving as a function of the environment statistics and barrier condition. Error bars represent SEM across animals.

(D) Consecutive failures before leaving split by subject and environment statistics, barrier versus no barrier.

with continued training this correlation decayed to zero (Figure 2F). Therefore, experienced mice, unlike naive animals, decide when to leave the foraging site in a manner consistent with inferring a hidden state rather than directly integrating rewards and failures.

As another way of seeing this, a stimulus-bound integration strategy would effectively weigh similarly each reward and failure with opposite signs (see Equation 1). Correct inference instead, given the structure of this task, requires that rewards are weighted nonlinearly (the first counting a lot and subsequent nothing) and differently from failures, which should add linearly. Indeed, in the trained mice, the effect of rewards and failures in shaping the behavior is qualitatively asymmetric in just this way, as can be seen by visualizing the probability of leaving as a function of both reward number and consecutive failures (Figure 2G).

Furthermore, stimulus-bound and inference-based models predict different interactions of rewards with preceding failures. Consider, for example, trials in which the animal receives a single reward: the later the reward, the smaller the value of the current site at the time of reward delivery. In the stimulus-bound model, the received reward value is simply added to the current value estimate, so the later in the train the reward arrives the lower the current value of the port (given that we are assuming only failures before this reward). Therefore, fewer subsequent failures will be tolerated before the animal leaves. On the other hand, in inference-based models a single reward resets the count of accumulated failures up to that point, and therefore the position of the reward (or equivalently the number of failures prior to the reward) has no consequence on subsequent behavior. To test these alternatives, we analyzed how the position of that reward influenced the overall number of failures

before leaving in a protocol with lower probability rewards ($p_{RWD} = 0.3$ and $p_{SW} = 0.3$) and found that the number of failures after the last reward did not decrease when it was preceded by more and more failures, on the contrary it slightly increased (Figure 2H, slope = 0.2 ± 0.06 , $n = 20$ mice), consistent with a resetting effect of reward as predicted by the inference-based model.

Accumulation of Evidence Is Tuned to Task Parameters

Having found that the foraging behavior of mice is consistent with the accumulation of evidence to infer a hidden world state, we asked whether this inference process is appropriately tuned to the statistics of the foraging environment, represented here by two parameters: reward probability p_{RWD} and state switch probability p_{SW} . Intuitively, if p_{RWD} is high, then a single failure is strong evidence in favor of a state switch, leading to a faster accumulation process. Similarly, if p_{SW} is high, then a failure also carries more evidence in favor of a state switch compared to if it is low (Figure 3A; see STAR Methods for a formal justification of this intuitive argument).

To test this, we trained a separate batch of mice on a set of three different foraging site statistics (Easy environment: $p_{RWD} = 0.9$ and $p_{SW} = 0.9$; Medium environment: $p_{RWD} = 0.9$ and $p_{SW} = 0.3$; Hard environment: $p_{RWD} = 0.3$ and $p_{SW} = 0.3$; see Figure 3B). Because changing the foraging environment's statistics can affect average reward rates (i.e., average number of rewards per trial), we adjusted the magnitude of individual rewards in order to equalize the amount of reward at a given site before state switch across conditions. As predicted normatively, mice increased the number of failed attempts they would tolerate as the state switching probability and the reward probability dropped (Figures 3C and 3D; difference in failed attempts after last reward in Easy-Medium = -1.61 ± 0.03 , difference Hard-Medium = 1.77 ± 0.04 , $n = 20$ mice, likelihood ratio test on $ConsecutiveFailures \sim 1 + Protocol + (1/MouseID)$ versus $ConsecutiveFailures \sim 1 + (1/MouseID)$; $p < 1e-10$).

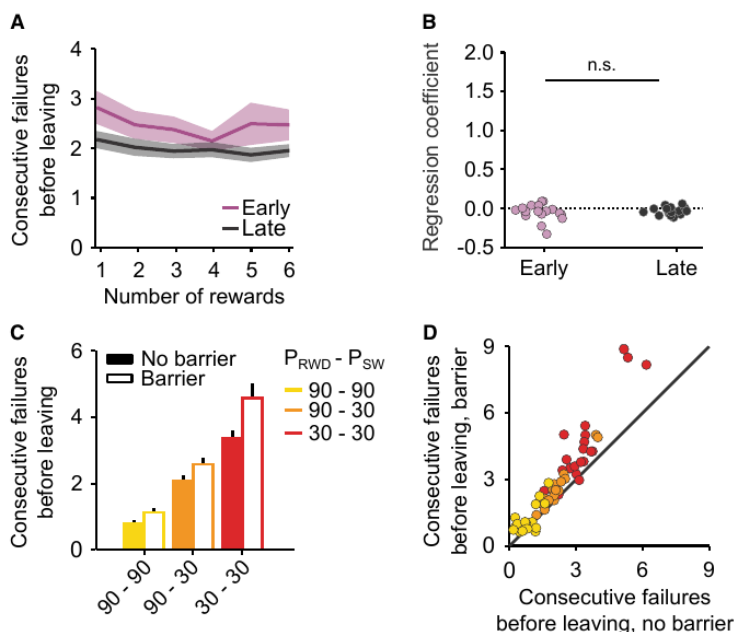


Figure 4. Humans Perform Optimal Inference and Tune Behavior to Task Parameters

(A) The number of rewards has little effect on the probability of leaving during both early (purple) and late (black) training. Solid line and shaded area represent mean and SEM across animals.

(B) Number of consecutive failures as a function of reward number for human in early versus late part of training. Unlike mice, humans learn the statistics of the environment extremely quickly: slope coefficient is similar (and around 0) in both early and late trials: likelihood ratio test on linear mixed-effect model $ConsecutiveFailures \sim 1 + RewardNumber + Early + RewardNumber \times Early + (1|SubjectID)$ versus a null model with no interaction: $p = 0.45$, $n = 20$ subjects.

(C) Consecutive failures before leaving as a function of the environment statistics and barrier condition. Error bars represent SEM across subjects.

(D) Consecutive failures before leaving split by subject and environment statistics, barrier versus no barrier. See also Video S2.

An important additional prediction of optimal decision theory in the context of a foraging task is that travel cost should modulate the threshold to leave a given foraging site. To test this, we increased the travel cost by placing a physical barrier between the two locations (travel time without barrier = 1.86 ± 0.13 s, $n = 20$ mice; travel time with barrier = 2.69 ± 0.13 s). Once again, the accumulation process was modulated consistently with the normative prediction, longer travel times resulting in a longer accumulation process and delayed leaving (Figures 3C and 3D, effect of barrier in number of failed attempts after last reward = 0.42 ± 0.03 , $n = 20$ mice, likelihood ratio test on $ConsecutiveFailures \sim 1 + Protocol + Barrier + (1|MouseID)$ versus $Consecutive Failures \sim 1 + Protocol + (1|MouseID)$: $p < 1e-10$).

Humans Perform Inference and Tune Behavior to Task Parameters

To test whether our findings were valid across species, we developed a translation of our behavioral assay for human subjects, in the form of a video game, where players would drag a character from one side of a touch screen to the other and tap to achieve points. The statistics of the video game (P_{RWD} and P_{SW}) were the same as those used in the rodent task.

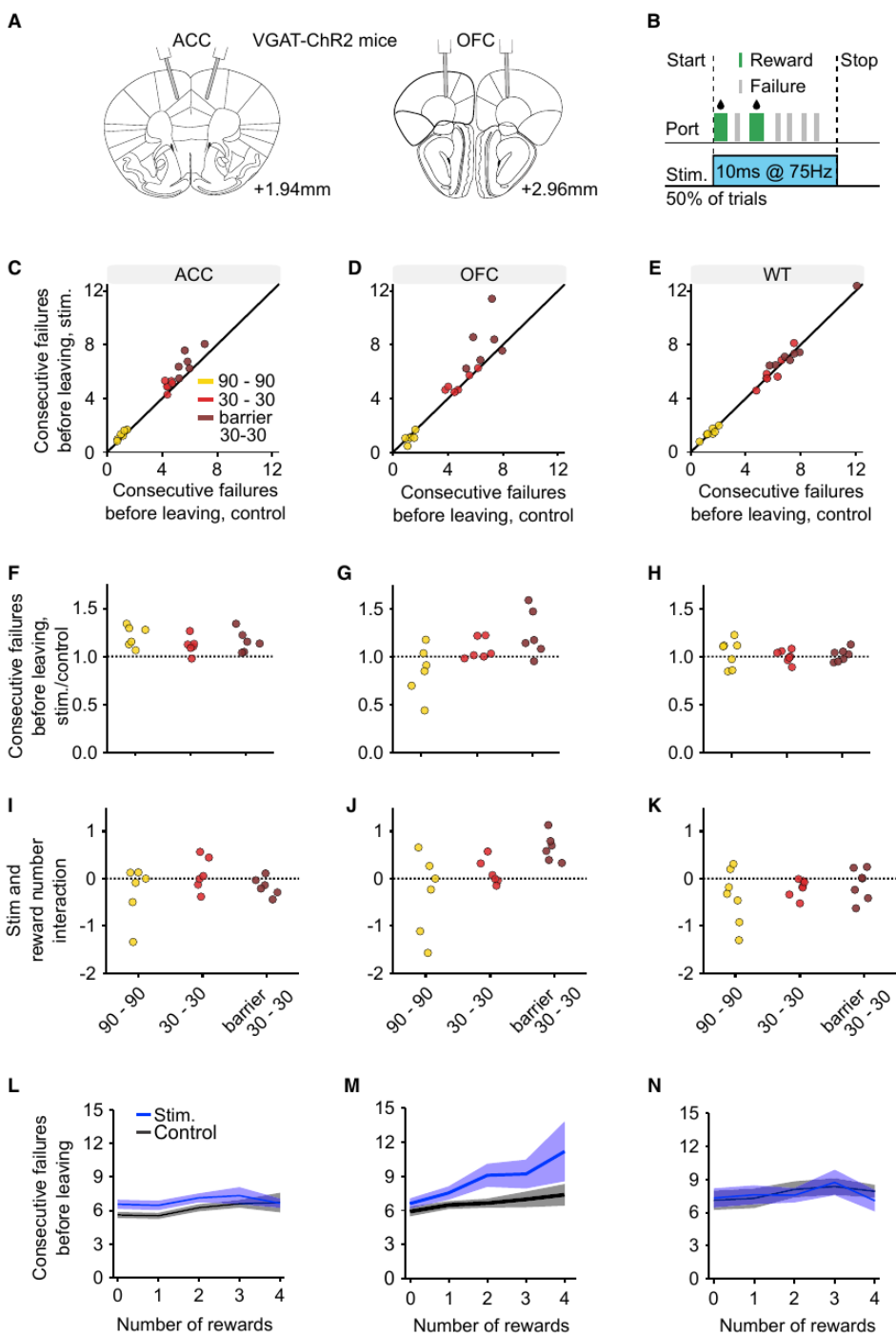
In humans, we again observed hallmarks of inference-based foraging: the number of rewards had little to no effect on the behavior (Figures 4A and 4B), similar to the behavior of the trained mice. Unlike mice, however, humans needed almost no training to learn this strategy, displaying it from the first session.

Analogously to their rodent counterparts, human subjects modulated their behavior according to reward statistics as well as travel time (here affected by a manipulation in the character's

velocity) consistent with the normative predictions (Figures 4C and 4D, difference in failed attempts after last reward in Easy-Medium = -1.39 ± 0.03 , difference Hard-Medium = 1.48 ± 0.03 , $n = 20$ subjects, likelihood ratio test on $ConsecutiveFailures \sim 1 + Protocol + (1|SubjectID)$ versus $ConsecutiveFailures \sim 1 + (1|SubjectID)$: $p < 1e-10$, effect of barrier in number of failed attempts after last reward = 0.59 ± 0.02 , $n = 20$ subjects, likelihood ratio test on $ConsecutiveFailures \sim 1 + Protocol + Barrier + (1|SubjectID)$ versus $Consecutive Failures \sim 1 + Protocol + (1|SubjectID)$: $p < 1e-10$).

OFC, but Not ACC, Is Necessary for the Correct Inference Process

Finally, to study the brain mechanisms of inference in this task, we tested the involvement of different regions of prefrontal cortex by silencing them using optogenetic stimulation of inhibitory GABAergic interneurons in VGAT-ChR2 mice (mice expressing the excitatory opsin channelrhodopsin-2 in inhibitory GABAergic neurons). We examined 19 mice. Nine were bilaterally implanted with optic fibers (Table S1) in the ACC (Figures 5A and S2A), six of these mice were ChR2-expressing (HET) and three were control wild-type littermates (WT) implanted and stimulated in the same manner. Ten (six HET and four WT) were bilaterally implanted in the OFC (Figures 5A and S2B). Transient inactivation of ACC (3 mW power per fiber, 10 ms pulses at 75 Hz, during poking; triggered by the first poke in 50% of trials and maintained for 500 ms after each poke in the trial; Figure 5B) significantly increased the average number of consecutive failures before leaving (Figure 5C, effect of stimulation on consecutive failures after last reward = 0.48 ± 0.05 , $n = 6$ mice, likelihood ratio test on $ConsecutiveFailures \sim 1 + Protocol + Stimulation + (1|MouseID)$ versus $ConsecutiveFailures \sim 1 + Protocol + (1|MouseID)$: $p < 1e-10$). The same protocol applied to control



mice had no effect (Figure 5E, effect of stimulation = 0.003 ± 0.07 , $n = 7$ mice, likelihood ratio test: $p = 0.96$). More specifically, we found that ACC inactivation multiplicatively increased the number of consecutive failures before leaving, consistently across protocols and animals (Figure 5F, *Protocol* and *Stimulation* interact when predicting *ConsecutiveFailures*, likelihood ratio test: $p = 1.46e-6$, but not when predicting renormalized *ConsecutiveFailures*, likelihood ratio test: $p = 0.15$, $n = 6$ mice).

Transient inactivation of OFC also increased the average number of consecutive failures before leaving (Figure 5D; effect of stimulation on consecutive failures after last reward = 0.41 ± 0.08 , $n = 6$ mice, likelihood ratio test on *ConsecutiveFailures* $\sim 1 + \text{Protocol} + \text{Stimulation} + (1|\text{MouseID})$ versus *ConsecutiveFailures* $\sim 1 + \text{Protocol} + (1|\text{MouseID})$: $p = 5.38e-7$). However, unlike the case for ACC inactivation, it did so in a manner that was dependent on the statistics of the environment. That is, the direction of effect for OFC inactivation actually reversed between easy and difficult protocols (effect of stimulation in hard protocol = 1.52 ± 0.24 , effect of stimulation in easy protocol = -0.19 ± 0.05 , $n = 6$ mice). This suggests that these two brain areas are differentially involved in the task.

To further investigate the involvement of these prefrontal areas in the inference process and task space representation, we considered a key difference in prediction between the inference model and the simpler stimulus-bound model: the effect of rewards on behavior. As noted above, under normal conditions, rewards fully reset the accumulation process, so that leaving times are not affected by the number of previous rewards (Figure 1E). Strikingly, we found that OFC, but not ACC inactivation, disrupted this pattern: in OFC-inactivated trials, animals became sensitive to the number of rewards: the more rewards gained, the more delayed leaving decisions became (Figures 5I and 5J; for ACC, interaction effect of stimulation and reward number = -0.038 ± 0.07 , $n = 6$ mice, likelihood ratio test on

versus

$$\begin{aligned} \text{ConsecutiveFailures} \sim & 1 + \text{Protocol} + \text{Stimulation} \\ & + \text{Stimulation} \& \text{Protocol} \\ & + \text{RewardNumber} + (1|\text{MouseID}) \end{aligned}$$

$p = 0.58$; for OFC, interaction effect of stimulation and reward number = 0.36 ± 0.1 , $n = 6$ mice, likelihood ratio test: $p = 0.0003$; triple interaction term of stimulation, reward number, and fiber location = 0.47 ± 0.11 , likelihood ratio test: $p = 3.44e-5$). This pattern of behavior (illustrated in Figures 5L and 5M) is similar to the one observed in naive mice first introduced to this task (Figures 2D and 2E) and is indicative of a less effective stimulus-bound strategy. Thus, the OFC is crucial for behavioral strategies in foraging environments in which states are hidden and require inference based on noisy observations.

DISCUSSION

In this study, we developed a task in which subjects had to alternate between two foraging sites, only one of which was active at any given moment. The task embodied an important form of non-sensory uncertainty because the active port only delivered rewards with a certain probability. The task thus required subjects to infer whether each omitted reward was simply a stochastic failure or was instead an actual switch of state, offering us a way to directly test whether they have the ability to perform state inference. To solve this task optimally, subjects were essentially required to infer a hidden state of the world (i.e., which site is active) rather than directly assigning a value to each foraging site, as would be optimal, for example, in a matching task (Herrnstein, 1961; Sugrue et al., 2004). We found that both mice and humans displayed hallmarks of optimal, inference-based behavior, reaching very similar solutions.

Our analysis of the behavioral data, particularly the number of consecutive non-rewarded tries before leaving, revealed that leaving decisions agreed with normative predictions of an inference-based foraging strategy in four important ways: (1) the number of consecutive failures was positively correlated with the propensity to leave; (2) rewards had a resetting effect on the leaving decision process; (3) subjects were sensitive to quantitative changes in the statistics of the foraging site; and (4) subjects were sensitive to the travel cost. However, mice and humans differed in an important way: while it took around

$$\begin{aligned} \text{ConsecutiveFailures} \sim & 1 + \text{Protocol} + \text{Stimulation} \\ & + \text{Stimulation} \& \text{Protocol} \\ & + \text{RewardNumber} \\ & + \text{Stimulation} \& \text{RewardNumber} \\ & + (1|\text{MouseID}) \end{aligned}$$

Figure 5. OFC, but Not ACC, Is Necessary for Optimal Inference

(A) Scheme of the optic fiber placement.

(B) Bilateral photostimulation at 3 mW happened during nose-poking: it was triggered by the first poke in 50% of trials and lasted for 500 ms after the last poke in the trial.

(C–E) Consecutive failures before leaving split by environment statistics, barrier condition, and subject inactivation versus control trials, for ACC implanted heterozygotes (C), OFC implanted heterozygotes (D), and wild-types (E), respectively.

(F–H) Ratio of consecutive failures before leaving split in the same way as (C)–(E) for ACC implanted heterozygotes (F), OFC implanted heterozygotes (G), and wild-types (H), respectively. When predicting renormalized consecutive failures, *Stimulation* and *Protocol* interact for OFC implanted heterozygotes ($p < 1e-10$, $n = 6$ mice) but not for ACC implanted heterozygotes ($p = 0.15$, $n = 6$ mice) or wild-types ($p = 0.77$, $n = 7$ mice).

(I–K) An animal by animal quantification: the coefficient of the interaction term in *ConsecutiveFailures* $\sim 1 + \text{Stimulation} + \text{Reward number} + \text{Reward number} \& \text{Stimulation}$ for ACC implanted heterozygotes (I), OFC implanted heterozygotes (J), and wild-types (K).

(L–N) Number of consecutive failures as a function of reward number in the 30–30 barrier protocol for ACC implanted heterozygotes (L), OFC implanted heterozygotes (M), and wild-types (N). Solid lines and shaded areas represent mean and SEM across animals.

See also Figure S2 and Table S1.

6 days for rodents to understand the environment statistics and integrate trial history correctly, humans started displaying hallmarks of the optimal behavior already during the first session. This difference may be due to faster learning but could also reflect the ability to generalize prior structural knowledge relevant to the task.

The accumulation of evidence is considered a primary cognitive computation. Similarly to sensory-guided tasks, in which integration of sensory evidence over time is needed to “average out” stimulus noise (Brunton et al., 2013; Gold and Shadlen, 2007; Shadlen and Newsome, 2001), here too, repeated sampling is needed to determine which of two sites is currently rewarding. Specifically, each failure conveys ambiguous information, as it may be due to either an unlucky attempt at the rewarding site, or a guaranteed failure in the non-active site, and it is only by counting (integrating) the number of consecutive failures that a more accurate state estimation can be made. Our analysis of the leaving probability revealed that, much like in sensory-based tasks, subjects do integrate this information when deciding whether to stay or leave. Moreover, by changing reward and transition probabilities, we were able to precisely control the amount of information associated with each failure, and observed that subjects readily adapted their leaving decisions to these changing conditions, such that the lower the information content of each sample was, the more such samples were needed before leaving.

In the framework of reinforcement learning under uncertainty, given the entire task history—i.e., the sequence of rewards and failures at each port since the beginning of the task—the optimal agent needs to compute a low-dimensional state representation (Niv, 2019) that is most informative of future events. For a formal definition, see Sect. 17.3 in Sutton and Barto (2018). We considered two distinct algorithms for doing this. In the “inference-based” algorithm, we hypothesize that animals recover a meaningful state representation that allows them to take most advantage of the task structure. The current state is represented by the posterior probability of being at the active site given the task history, which in practice is a function of the number of consecutive failures since the last reward. In this algorithm, the final learned solution is optimal and independent of the learning rate used during training. Alternatively, in the “stimulus-bound” algorithm, the animal only uses observable states based on currently available perceptual information (Wilson et al., 2014). The entire task history is summarized by the current location of the animal (left or right site). In the hidden state task, this representation only allows for suboptimal stable solutions, i.e., policies that depend on the site, but not on the reward history.

The primary distinction between sub-optimal, stimulus-bound and optimal, inference-based strategies lies in the impact consecutive rewards have on leaving decisions. In a stimulus-bound behavior, which assigns values directly to the foraging site, the more consecutive rewards are gained at a given site, the higher the value of staying becomes, and consequently, leaving decisions tend to be delayed. In contrast, optimal inference in this task requires ignoring the number of consecutive rewards, because the delivery of a single reward is sufficient to know for certain which site is currently rewarding. As shown in Figure 2, we found that, although initial behavior appeared to be sensory

bound, after learning, subjects’ leaving decisions became independent of the number of rewards, consistent with an inference-based approach to leaving decisions. We presume that the change in behavior (between stimulus-bound and inference-based decisions) over the course of training reflects learning, but not that the change in performance necessarily reflects a change in learning rate. What we posit is that two different behavioral controllers, one stimulus-bound and one inference-based, exist, which might correspond to a striatal system and a prefrontal system respectively. Over the course of training, the inference-based controller learns the structure of the task—the correct state representation—through a slow process. In parallel, the inference-based controller’s contribution to behavioral choices is increased over training. This scheme is similar to what was proposed by Daw et al. (2005). Alternatively, the change in behavior over time could be accomplished by meta-learning of hyper-parameters: a more complex stimulus-bound agent could keep track of two different learning rates, one for reward and one for failures. With training, the agent would learn that the optimal reward learning rate is one (complete reset) whereas the failure learning rate is adjusted over time to account for different protocols. Even though this algorithm is distinct, it still requires the ability to adjust a failure learning rate in such a way that it is big for informative failures (in easy protocols) and small for uninformative ones (in harder protocols). Consequently, it results in a computation analogous to the inference model.

Recent accounts (Niv, 2019; Schuck et al., 2016; Stalnaker et al., 2015; Wilson et al., 2014) proposed that the OFC is crucial for accurate state representations, particularly when states are hidden (that is, not explicitly given by the presence of a sensory cue, for example) and have to be inferred from fuzzy evidence. Our findings mesh well with this theory, because we found that OFC, but not ACC, inhibition disrupted inference-based behavior. Unlike under control conditions, in which the number of failures before leaving was independent of the number of previously gained rewards, inhibiting the OFC resulted in mice performing more failures when experiencing large amounts of reward. This latter pattern is consistent with a stimulus-bound strategy, and suggests the possibility that the stimulus-bound strategy serves as a default behavioral approach, and is suppressed by the OFC when inference-based behavior is required. The observation that naive mice behave very similarly to OFC-inactivated mice supports this idea. Several specific computational roles of OFC could account for this effect. OFC could be encoding the representation directly or be necessary to access or update such representation. Alternatively, the representation could be still available, even when the OFC is inactivated, but the region would be responsible for computing the posterior of the hidden states given the representation.

The ACC has been implicated, in Pavlovian and operant tasks, as a potential candidate for the implementation of integration-to-threshold models. In Kawai et al. (2015), the authors observe neurons in the primate ACC whose firing rate scales with the number of consecutive negative outcomes in a Pavlovian task. In the setting of an operant task, Sarafyazd and Jazayeri (2019) showed that such negative outcome accumulation is modulated by the error type (the more surprising the error, the stronger the response) and that microstimulation of ACC accelerates the

detection of a context switch. From the foraging perspective, Hayden et al. (2011) reported cells in the ACC encoding the value of a depleting option. However, ACC inactivation in our task, unlike OFC inactivation, had only a modulatory effect on behavior: we did not observe qualitative changes in the strategy of the animals, but only an overall tendency to stay longer at the current port, which interacted multiplicatively both with the task statistics and with increased travel times. The potential activation of neurons in regions immediately adjacent to ACC (e.g., prelimbic cortex or secondary motor cortex) is possible (Figure S2C). However, the areas targeted in the two experiments (ACC and OFC) are considerably further apart (Figure S2C). Because we observed a double-dissociation of effects it is unlikely that the fields of neurons activated across these two experiments were substantially overlapping. Although caution is required in comparing primate and rodent ACC, given the reported anatomical (Uylings and van Eden, 1990) and functional differences (Narayanan et al., 2013; Seamans et al., 2008), our results seem most compatible with the idea that the ACC encodes the value of alternative options, as proposed in Kolling et al. (2016), while not having a primary role in the computations required for the state inference process.

By developing a human video game and a rodent task requiring the same underlying computation to be solved, we could compare computational and cognitive processes across species. From a theoretical standpoint, this strengthens the generality of those results that held true for the two species, such as the ability to infer the hidden structure of the environment and to tune behavior to environmental statistics. From a practical standpoint, the hidden state foraging task makes it possible to use rodent experimentation to more closely guide human clinical research into the mechanisms of manipulations (e.g., drugs) or conditions (e.g., depression) that may affect processes such as state inference.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Mice
 - Human participants
- **METHOD DETAILS**
 - Mice behavioral apparatus
 - Human video game task
 - Optogenetic stimulation
 - Surgical procedures
 - Histology and microscopy
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Statistical analysis
 - Task design
 - Relevant statistics in the task
 - Possible task space representations
 - Stimulus-bound learning
 - Inference-based learning

- From value to decision
- Computing the likelihood ratio
- **DATA AND CODE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.neuron.2020.01.017>.

ACKNOWLEDGMENTS

We would like to thank Shira Lottem for her help with the graphic art of the human task. This work was supported by the European Research Council (Advanced Investigator Grant 671251 to Z.F.M.) and Champalimaud Foundation (Z.F.M.).

AUTHOR CONTRIBUTIONS

P.V., E.L., D.S., and Z.F.M. designed the experiments and analyses. P.V., D.S., B.G., I.T., T.Q., and M.N.O.L. conducted the experiments. P.V. and E.L. analyzed the data. P.V. wrote the original draft. E.L., D.S., and Z.F.M. reviewed and edited the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 14, 2019

Revised: October 24, 2019

Accepted: January 14, 2020

Published: February 11, 2020

REFERENCES

- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Rev.* 59, 65–98.
- Boyer, X., Friedman, N., and Koller, D. (1999). Discovering the Hidden Structure of Complex Dynamic Systems. *arXiv*, arXiv:1301.6683 <https://arxiv.org/abs/1301.6683>.
- Braun, D.A., Mehring, C., and Wolpert, D.M. (2010). Structure learning in action. *Behav. Brain Res.* 206, 157–165.
- 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.
- Charnov, E.L. (1976). Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* 9, 129–136.
- Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711.
- Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215.
- Eshel, N., Tian, J., Bukwich, M., and Uchida, N. (2016). Dopamine neurons share common response function for reward prediction error. *Nat. Neurosci.* 19, 479–486.
- Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574.
- Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nat. Neurosci.* 14, 933–939.
- Herrnstein, R.J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* 4, 267–272.
- Kawai, T., Yamada, H., Sato, N., Takada, M., and Matsumoto, M. (2015). Roles of the Lateral Habenula and Anterior Cingulate Cortex in Negative Outcome Monitoring and Behavioral Adjustment in Nonhuman Primates. *Neuron* 88, 792–804.

- Kolling, N., Behrens, T.E.J., Mars, R.B., and Rushworth, M.F.S. (2012). Neural mechanisms of foraging. *Science* 336, 95–98.
- Kolling, N., Wittmann, M., and Rushworth, M.F.S. (2014). Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* 81, 1190–1202.
- Kolling, N., Wittmann, M.K., Behrens, T.E.J., Boorman, E.D., Mars, R.B., and Rushworth, M.F.S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* 19, 1280–1285.
- Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176.
- Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., Lohuis, M.O., and Mainen, Z.F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nat. Commun.* 9, 1000.
- McLean, R.A., Sanders, W.L., and Stroup, W.W. (1991). A Unified Approach to Mixed Linear Models. *The American Statistician* 45, 54–64.
- Narayanan, N.S., Cavanagh, J.F., Frank, M.J., and Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. *Nat. Neurosci.* 16, 1888–1895.
- Niv, Y. (2019). Learning task-state representations. *Nat. Neurosci.* 22, 1544–1553.
- Niv, Y., Daniel, R., Geana, A., Gershman, S.J., Leong, Y.C., Radulescu, A., and Wilson, R.C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *J. Neurosci.* 35, 8145–8157.
- Pearl, J. (1991). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. *J. Philos.* 88, 434–437.
- Rudebeck, P.H., Saunders, R.C., Prescott, A.T., Chau, L.S., and Murray, E.A. (2013). Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nat. Neurosci.* 16, 1140–1145.
- Sarafyzd, M., and Jazayeri, M. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. *Science* 364, eaav8911.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.
- Schuck, N.W., Cai, M.B., Wilson, R.C., and Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron* 91, 1402–1412.
- Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.
- Seamans, J.K., Lapish, C.C., and Durstewitz, D. (2008). Comparing the prefrontal cortex of rats and primates: insights from electrophysiology. *Neurotox. Res.* 14, 249–262.
- 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.
- Shima, K., and Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282, 1335–1338.
- Stalnaker, T.A., Cooch, N.K., and Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. *Nat. Neurosci.* 18, 620–627.
- Starkweather, C.K., Gershman, S.J., and Uchida, N. (2018). The Medial Prefrontal Cortex Shapes Dopamine Reward Prediction Errors under State Uncertainty. *Neuron* 98, 616–629.
- 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.
- Sutton, R.S., and Barto, A.G. (2018). Reinforcement Learning: An Introduction (MIT Press).
- Tervo, D.G.R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., and Karpova, A.Y. (2014). Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. *Cell* 159, 21–32.
- Uylings, H.B.M., and van Eden, C.G. (1990). Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. *Prog. Brain Res.* 85, 31–62.
- Wilkinson, G.N., and Rogers, C.E. (1973). Symbolic description of factorial models for analysis of variance. *J. R. Stat. Soc. C-Appl.* 22, 392–399.
- Wilks, S.S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Ann. Math. Statist.* 9, 60–62.
- Williams, Z.M., Bush, G., Rauch, S.L., Cosgrove, G.R., and Eskandar, E.N. (2004). Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat. Neurosci.* 7, 1370–1375.
- 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.
- Bates, D., José Bayoán Santiago Calderón, Kleinschmidt, D., Kelman, T., Babayan, S., Mogensen, P.K., Piibeleht, M., Bouchet-Valat, M., Hatherly, M., Saba, E., et al. (2019). dmbates/MixedModels.jl: Avoid fallback to generic_matmul.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
DAPI	SIGMA ALDRICH	Cat#D9542; RRID:AB_2801570
Experimental Models: Organisms/Strains		
C57BL/6NCrI	Charles River Laboratories	strain code: 027; RRID:IMSR_CRL:475
Dat-Cre	Jackson Laboratory	stock number: 006660; RRID:IMSR_JAX:006660
Gad-Cre	Jackson Laboratory	stock number: 010802; RRID:IMSR_JAX:010802
VGAT-ChR2	Jackson Laboratory	stock number: 014548; RRID:IMSR_JAX:014548
Fl12-Cre	Mutant Mouse Regional Resource Centers	stock number: 017262-UCD; RRID:MMRRC_017262-UCD
Sert-Cre mouse line 61	Mutant Mouse Regional Resource Centers	stock number: 017260-UCD; RRID:MMRRC_017260-UCD
Software and Algorithms		
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
Construct 2	Scirra Ltd.	https://www.scirra.com/construct2
Julia language	Bezanson et al., 2017	https://julialang.org/
MixedModels.jl	Bates et al., 2019	https://github.com/JuliaStats/MixedModels.jl
Other		
Arduino Mega 2560 r3	Arduino	A000067
Pokes detector and valve controller	Champalimaud Hardware Platform	Mice poke simple v1.1
Arduino ports interface	Champalimaud Hardware Platform	Arduino baseboard v2.2

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for reagents may be directed to, and will be fulfilled by the Lead Contact, Zachary Mainen (zmainen@neuro.fchampalimaud.org). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

Fifty-seven adult male C57BL/6 mice were used in this study. For the inference-based versus stimulus-bound behavior experiment (Figure 2) 18 C57BL/6NCrI wild-type mice of two months age were used. For the protocols manipulation experiment (Figure 3), 20 wild-type animal from different genetic backgrounds (8 Dat-Cre ; 5 Gad2-Cre; 5 Sert-Cre; 1 VGAT-ChR2; 1 F12-Cre) of 6-8 months age were used, in order to reduce animal usage. For inactivation of anterior cingulate or orbitofrontal cortices (Figure 5), 12 VGAT-ChR2 and 7 wild-type littermates were used. Mice genotypes were determined based on PCR and further verified using histological inspection of YFP expression which led to the exclusion of a single ACC implanted animal from further analysis (see Figure S2d, e). The C57BL/6NCrI line was obtained from the Charles river laboratories, breeders were ordered and bred in-house for a maximum of 4 generations or 2 years (strain code: 027). The Dat-Cre mouse line was obtained from the Jackson laboratory (stock number: 006660). The Gad2-Cre was obtained from the Jackson laboratory (stock number: 010802). The Sert-Cre mouse line 61 was obtained from the Mutant Mouse Regional Resource Centers (stock number: 017260-UCD). The VGAT-ChR2 mouse line 8 was obtained from the Jackson laboratory (stock number: 014548). The Fl12-Cre mouse line was obtained from the Mutant Mouse Regional Resource Centers (stock number: 017262-UCD). All experimental procedures were approved and performed in accordance with the Champalimaud Centre for the Unknown Ethics Committee guidelines and by the Portuguese Veterinary General Board(Direcao-Geral de Veterinaria, approval 0421/000/000/2016). The mice were kept under a normal 12 h light/dark cycle, and training, as well as testing, occurred during the light period. Before testing or after surgeries, for the inactivation experiments, mice were single-housed. During training and testing the mice were water deprived, and water was available to them only during task performance. Food was freely accessible to the mice in their home cages. Extra water was provided if needed to ensure that mice maintain no less than 80% of their original weight. For the protocols manipulation experiment behavioral training lasted 12 sessions, once per day, followed by 2 days of rest at

the end of which we commenced testing. During training mice were exposed to the 3 different protocols (Easy: $p_{RWD} = 0.9$ and $p_{SW} = 0.9$; Medium: $p_{RWD} = 0.9$ and $p_{SW} = 0.3$; Hard: $p_{RWD} = 0.3$ and $p_{SW} = 0.3$) for 4 consecutive days (1 day of adaptation and 3 of testing) before transitioning to the next environment. During testing, mice performed 1 session per day, 6 or 7 days a week. In protocols manipulation experiments, the sequence of protocols was counterbalanced across 2 groups of 10 mice (Group A: Hard, Medium, Easy; GroupB: Easy, Medium, Hard). In our analyses we considered 50 poke bouts per session after the first 10 during testing days and excluded poke bouts with no rewards.

Human participants

20 right handed healthy adults of Portuguese nationality (10 female and 10 male; 22 to 31 years of age), with no history of psychiatric diagnosis or prescribed drugs in the last 6 months, participated in this study. All participants gave written informed consent, and the study was conducted in accordance with the guidelines of the local ethics committee. The task consisted of 2 sessions of 1 hour, performed in different days with 2 to 10 days in between sessions. Each session consisted of 4 blocks with different protocols, and 10 minutes break after the second block. The sequence of protocols consisted of a block (Medium environment) followed by a short break (2 minutes), then a second block (Easy or Hard environment) followed by a long break (10 minutes), then a third block (Medium environment) followed by a short break (2 minutes) and a final block (Hard or Easy environment). The sequence of environments during testing was counterbalanced across 2 groups as described in mice experiments. In our analyses we considered all tapping bouts after the first 10 and excluded bouts with no rewards.

METHOD DETAILS

Mice behavioral apparatus

The behavioral apparatus for the task was adapted from the design developed by Zachary F. Mainen and Matt Recchia (Island motion corporation, Nesconset, NY), originally developed for rat behavior. The behavioral box (15 × 12 × 18 cm, model 003102.0001, Island motion corporation), contained 3 front walls (135-degree angle between the center and the side walls) with 2 nose-poke ports attached to the left and right front walls. For the inference-based versus stimulus-bound behavior experiment (Figure 2), we used a custom-made acrylic replicate of the box (15 × 16 × 20 cm). Each port was equipped with infrared emitter/sensor pairs to report the times of port entry and exit (model 007120.0002, Island motion corporation). A nose-poke was considered valid if the infrared beam was broken for at least 100 ms. Water valves (LHDA1233115H, The Lee Company, Westbrook, CT) were calibrated to deliver a drop of 6 μ l water for rewarded pokes in Easy and Hard environments and 2 μ l of water in Medium environment: the reward size was adjusted to keep the reward amount per correct trial constant. The average number of rewarded attempts per correct trial is p_{RWD}/p_{SW} , that is to say 1 in the easy and hard protocol (reward magnitude = 6 μ l, amount of water per correct trial = 6 μ l) and 3 in the medium protocol (reward magnitude = 2 μ l, amount of water per correct trial = 6 μ l). In optogenetic experiments, all protocols had an average of one reward per trial, but the reward size was kept at 4 μ l to increase the trial number. In optogenetic experiments, blue LEDs were placed in the box ceiling and in all the ports to deliver a masking light. All signals from sensors were processed by Arduino Mega 2560 microcontroller board (Arduino, Somerville, US) and output from the Arduino Mega 2560 microcontroller board was implemented to control water and light delivery. Arduino Mega 2560 microcontroller was connected to the sensors and controllers through an Arduino Mega 2560 adaptor board developed by the Champalimaud Foundation Scientific Hardware Platform. An example behavioral video is available in the supplemental information.

Human video game task

Human subjects played a video game on a touchscreen device, with analogous features to the rodent behavioral assay. In the game, subjects receive verbal instructions on how to control a character - a "witch" - on its quest to find and defeat an enemy that hides behind a castle. The witch must walk along the wall of a castle, shooting either the left or the right edge of this wall in search of the enemy that hides, at any given moment, in one of these two edges. The game obeys the same statistics as the rodent task: hitting the enemy is analogous to a water reward, the current location of the enemy corresponds to the active site, and every shot at the active site hits the enemy with probability p_{RWD} . Moving between the two sides of the wall has an associated cost (travel cost) that can also be manipulated with the appearance of rougher terrain (analogous to the physical barrier) that diminishes the traveling speed. As in the mouse case, reward size was manipulated to keep the average reward per correct bout constant (3 points). The game ended either when subjects collected 280 points or when a time limit of 20 minutes was exceeded.

Different environments had minor changes in the background images between them - for the medium protocol since it was experienced twice per session, two different backgrounds were used. After the player transitioned to a different site, the enemy was displayed to cue whether the transition had been correct or if instead, the player had to return to the previous site.

The human task was made using custom software developed using the game engine Construct2 (Scirra Ltd., Studio 117, The Light Bulb 1 Filament Walk Wandsworth, London, UK). Graphics were made by Shira Lottem and Tiago Quendera using Inkscape: Open Source Scalable Vector Graphics Editor. Audio assets were made by Tiago Quendera using Audacity(R) except for the Wilhelm Scream (Wikimedia Commons).

An example video (not from an experimental subject) showing the different environments is provided in the supplemental material. The task, open-source code and all assets are available at <https://github.com/quendera/human-foraging>.

Optogenetic stimulation

In order to optically stimulate ChR2 expressing VGAT-expressing GABAergic interneurons we used blue light from a 473 nm laser (LRS-0473-PFF-00800-03, Laserglow Technologies, Toronto, CA or DHOM-M-473-200, UltraLasers, Inc., Newmarket, CA) that was controlled by an acousto-optical modulator (AOM; MTS110-A1-VIS or MTS110-A3- VIS, AA optoelectronic, Orsay, FR) to deliver light 10 ms pulses of light at 75 Hz, connected to Arduino Mega 2560 microcontroller board (Arduino, Somerville, US). Light exiting the AOM was focused into an optical fiber patch cord (200 μ m, 0.22 NA, Doric lenses Inc, 357 rue Franquet, Quebec, Quebec, CA), connected to a second fiber patch cord through a rotary joint (FRJ 1x1, Doric lenses), which was then connected to the chronically implanted optic fiber cannula (MFC_200/230-0.48_3mm_ZF1.25(G)_FLT; Doric lenses Inc, 357 rue Franquet, Quebec, Quebec, CA). We estimated an average 15% loss of light power between the patch cord tip and the optic fiber cannula before surgery. In order to deliver light at 3 mW power, previously to each experiment day, the laser power at the tip of the patch cord was adjusted to 3.6 mW, to account for the estimated power loss. To test each protocol, we habituated animals to the new protocol for two days, then stimulated during six consecutive days. Stimulation was delivered on 50% of trials and started with the first valid nose-poke (that is to say after the infrared beam was broken for at least 100 ms). Stimulation ended if the animal did not nose-poke for 500 ms but would restart in case of another valid nose-poke on the same site.

Surgical procedures

Animals were anesthetized with isoflurane (4% induction and 0.5 - 1% for maintenance) and placed in a motorized computer-controlled Stoelting stereotaxic instrument with mouse brain atlas integration and real-time visualization of the surgery probe in the atlas space (Neurostar, Sindelfingen, Germany; <https://www.neurostar.de>). Antibiotic (Enrofloxacin, 2.5-5 mg/Kg, S.C.) and pain killer (Buprenorphine, 0.1 mg/Kg, S.C.) and local anesthesia over the scalp (0.2 ml, 2% Lidocaine, S.C.) were administered before incising the scalp. Target coordinates were 1.9 mm A.P., $\pm$ 0.5 mm M.L., 1.75 mm D.V. for ACC and 2.9 mm A.P., $\pm$ 1.25 mm M.L., 1.8 mm D.V. for OFC. Two craniotomies were performed above the target's coordinates for OFC implants. For ACC implants fiber were implanted over the target with an angle of $\pm$ 16° on the ML axis to avoid damage to the superior sagittal sinus, and two craniotomies were performed at coordinates 1.9 mm A.P., $\pm$ 1 mm M.L. An optical fiber (200 μ m core diameter, 0.48 NA, 510 mm) housed inside a connectorized implant (M3, Doric lenses, Quebec, Canada) was lowered into the brain (0 degree angle for OFC and 11 degree angle for ACC), through the craniotomy as the viral injection, and positioned 10 μ m above the target. The implant was cemented to the skull using Super Bond C&B (Morita, Kyoto, Japan) and once dried covered with black dental cement acrylic (Pi-Ku-Plast HP 36, Bredent, Senden, Germany). The skin was stitched at the front and rear of the implant. Gentamicin (48760, Sigma-Aldrich, St. Louis, MO) was topically applied around the implant. Mice were monitored until recovery from the surgery and returned to their home cage where they were housed individually. Gentamicin (48760, Sigma-Aldrich, St. Louis, MO) was topically applied around the implant. Behavioral testing started at least 1 week after surgery to allow for recovery.

Histology and microscopy

For histological analysis mice were perfused transcardially with 4% paraformaldehyde (PFA) in phosphate buffer solution (PBS). After removing the brain they were left for 24 hours in 4% PFA solution in PBS, then transferred in 0.1% sodium azide solution in PBS. Brains were sliced in 50 μ m coronal sections on a vibratome (Leica VT 1000 S), collected in wells maintaining the anterior-posterior order, and finally mounted on microscope slides (Thermo scientific, superfrost plus), with mowiol.

Fluorescent images were acquired with an automated slide scanner (AxioScan Z1) equipped with a 10x, 0.45 NA PlanApochromat objective and a Hamamatsu OrcaFlash camera. Use of the appropriate filter combination allowed for DAPI and EYFP acquisition (Beam Splitter: 395, excitation: 330-375, emission: 430-470, and Beam splitter: 498, excitation: 453-485, emission: 507-546 respectively).

Optic cannula placement was determined using coronal sections of the prefrontal cortex through which the fiber tract was visible. We determined the position by locating the section with the broadest base of the cannula tract and comparing the DAPI staining with the Allen Mouse Brain Atlas (Lein et al., 2007) (Figure S2; Table S1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

The statistical analysis, which can be found in the Results section and figure legends, was performed using mixed effect models (McLean et al., 1991), in particular the Julia (Bezanson et al., 2017) implementation MixedModels.jl (Bates et al., 2019). For each mixed model, we report the maximum likelihood estimate of the coefficient of interest $\pm$ the standard error of the estimate. Our N is the number of subjects: as different experiments had a potentially different number of subjects, we report it after every statistical test. We fitted models with a random intercept (depending on subject identity) and compared nested models using a likelihood ratio test: in particular we used a chi-square test on the difference of the deviance of the two nested models, using as many degrees of

freedom as the difference between the number of degrees of freedom of the two nested models (Wilks, 1938). That is to say, given two models m and n where n is a special case of m :

$$p = 1 - \text{cdf}\left(\chi^2_{\text{dof}(m) - \text{dof}(n)}, \text{deviance}(n) - \text{deviance}(m)\right)$$

When the p value is too small, we do not report the value but simply write $p < 1\text{e-}10$, which is floating point notation for $p < 10^{-10}$. To describe mixed models we will use Wilkinson notation (Wilkinson and Rogers, 1973), with $|$ denoting random effects and $\&$ denoting interaction terms. For example the formula:

$$\text{ConsecutiveFailures} \sim 1 + \text{Protocol} + \text{Stimulation} + (1|\text{MouseID})$$

uses as predictor for the number of consecutive failures after last reward a constant intercept, a coefficient for each protocol different than the medium protocol (which we consider as baseline), a coefficient for stimulation and a random intercept across mice. The formula

$$\text{ConsecutiveFailures} \sim 1 + \text{Protocol} + \text{Stimulation} + \text{Protocol} \& \text{Stimulation} + (1|\text{MouseID})$$

would also allow for an interaction term between protocol and stimulation.

We did not test whether the data met the assumptions of the statistical methods used.

Task design

We designed a probabilistic foraging task for parallel use in mice and humans. Subjects sought rewards (water or points, respectively), by actively probing a foraging site (nose-poking or screen-tapping, respectively). Each site could be in one of two states, active or inactive. Each try in the active state yielded reward with probability p_{RWD} , and could cause the site to switch to the inactive state with probability p_{SW} . This required the subjects to travel to a second, fresh, site at some distance and bear a travel cost. Subjects were therefore tasked with inferring a hidden state (active or inactive) through a stochastic sequence of observations (rewards and failures).

Relevant statistics in the task

After a rewarded attempt, the subject could be sure to be in the correct location: ambiguity comes from failures, as it was possible that the target was correct but the subject was being unlucky. The more unsuccessful attempts, the higher the probability of a transition having occurred. Accumulated evidence in favor of a switch is a monotonically increasing function of the task parameters p_{RWD} and p_{SW} : the higher the reward probability, the more informative a failure is. Trivially, the higher the switch probability, the more likely the switch.

Possible task space representations

When analyzing our task, we consider two possible state representations. One is simpler and analogous to traditional approaches to modeling n -armed bandit tasks: the state corresponds to the current location of the subject (i.e., one of the two reward sites). The value of the two sites changes over time, yet the animals may be able to track this change with fast model-free learning. A second, more principled but more abstract approach, postulates that the subjects tries to infer the optimal state representation, i.e., the probability that their current location is rewarding. In this model, there is no longer any need for fast online learning as the task representation is stable. The computation happening in real time is the inference process to compute this probability. To account for variability in the behavior, we allow both decision noise, distributed according to the soft-max rule, as well as inference noise (the inference process may be suboptimal). We will refer to these two learning paradigms as stimulus-bound learning and inference-based learning respectively. It is important to note that, given the richer state representation, inference-based learning is the optimal way to solve the task and clearly outperforms simpler heuristics such as stimulus-bound learning.

Stimulus-bound learning

In stimulus-bound learning, we first define the relative value V as the difference of the value of the leftport and the value of the right port:

$$V = V_{\text{LEFT}} - V_{\text{RIGHT}} \quad (\text{S1})$$

We defined two auxiliary variables: a reward variable r indicating the outcome of each reward attempts, i.e., 1 for a reward and 0 for a failure, and a site variable s , indicating the current site, 1 for left and -1 for right.

We can define a signed outcome o of each reward attempt, which is:

$$o = r \cdot s \quad (\text{S2})$$

that is to say, 1 for a reward on the left, -1 for a reward on the right and 0 for an omission.

For any attempt, we can then update the relative value using the signed outcome and some discount parameter γ

$$V_{t+1} = (1 - \gamma)V_t + \gamma o_{t+1} \quad (\text{S3})$$

Which admits an explicit solution:

$$V_t = (1 - \gamma)^t \cdot V_0 + \gamma \cdot \sum_{i=1}^t (1 - \gamma)^{t-i} \cdot o_i \quad (S4)$$

The probability of staying is a monotonically increasing function of the value of staying, so that rewards should make the animal more likely to stay and omissions more likely to leave, in a symmetric way.

Inference-based learning

We first derive recursive formulas to compute the probability that the current site is not rewarding as a function of the sequence of successful and failed reward attempts performed by the subject. In this model, the subject would compute the relative value as a function of the probability of the left (or right) site being active given the task history ($r_1, \dots, r_t$ represent the outcomes of the various attempts and $s_1, \dots, s_t$ the site of each attempt):

$$V_t = p_{RWD} (P(\text{Lef } t\text{Active} | r_1, s_1, \dots, r_t, s_t) - P(\text{RightActive} | r_1, s_1, \dots, r_t, s_t)) \quad (S5)$$

From value to decision

We have now defined to different ways to compute the relative value of left versus right, one directly based on reward accumulation, and one based on evidence accumulation. To define a behavior from this relative value, we need to consider two more parameters. First of all we need a bias term $\square$: as the two foraging sites are far apart, subjects should prefer to repeat site rather than alternate, to avoid the travel cost. Then we need a “inverse temperature” parameter β to describe how deterministic the animal is (with a very high β the animal would almost always choose the option with greater value, whereas with $\beta = 0$ the animal would choose randomly). We can then use the soft-max rule to generate behavior:

$$P(\text{NextLef } t | V, s) = \sigma(\beta(V + s \cdot T)) \quad (S6)$$

where s represents the current site (1 for left and -1 for right).

In the simulations we will use the same softmax rule to simulate behavior: the difference between stimulus-bound and inference-based learning derives from the different procedures used to compute the relative value.

Computing the likelihood ratio

In inference-based learning we defined the relative value as a function of the relative difference:

$$P(\text{Lef } t\text{Active} | r_1, s_1, \dots, r_t, s_t) - P(\text{RightActive} | r_1, s_1, \dots, r_t, s_t) \quad (S7)$$

This quantity can be computed recursively. To do so we will need an auxiliary variable. We define R_t as the probability ratio that the current site is active or inactive given task history:

$$R_t = \frac{P(\text{Inactive} | r_1, s_1, \dots, r_t, s_t)}{P(\text{Active} | r_1, s_1, \dots, r_t, s_t)} \quad (S8)$$

From R_t we can compute V_t as follows:

$$V_t = p_{RWD} \cdot \left(1 - \frac{2}{R_t^{s_t} + 1} \right) \quad (S9)$$

Rather than computing V_t recursively directly, we notice that R_t respects a simple recursive equation (likelihood ratio update equation):

$$R_{t+1} = \left(\frac{R_t + p_{SW}}{1 - p_{SW}} \right)^{s_t s_{t+1}} \cdot \frac{P(r_{t+1} | \text{Inactive})}{P(r_{t+1} | \text{Active})} \quad (S10)$$

where p_{SW} represents the probability of switching from active to inactive state.

The term $(R_t + p_{SW} / 1 - p_{SW})$ is the ratio between the following two equations:

$$P(\text{NextInactive}) = P(\text{Inactive}) + p_{SW} P(\text{Active}) \quad (S11)$$

$$P(\text{NextActive}) = (1 - p_{SW}) \cdot P(\text{Active})$$

The exponent $s_t s_{t+1}$ simply means that the probability ratio $(P(\text{Inactive}) / P(\text{Active}))$ inverts when the subject changes site. Finally the term $(P(r_{t+1} | \text{Inactive}) / P(r_{t+1} | \text{Active}))$ represents the new evidence acquired with the outcome of attempt $t + 1$.

In the case of a reward, $P(r_{t+1}|Inactive) = 0$, so:

$$R_{t+1} = 0 \quad (S12)$$

If r_{t+1} is a failure, then $P(r_{t+1}|Inactive) = 1$ whereas $P(r_{t+1}|Active) = 1 - p_{RWD}$, so the likelihood ratio update equation simplifies to:

$$R_{t+1} = \left(\frac{R_t + p_{SW}}{1 - p_{SW}} \right)^{s_t s_{t+1}} \cdot \frac{1}{1 - p_{RWD}} \quad (S13)$$

Having established that R_t resets to 0 with a reward, we can analyze the most interesting case for a probability computation: a sequence of attempts on the same site (let us say $s_1, \dots, s_t = 1$) where the first attempt is rewarded (thus resetting the probability) and the following are not.

As the first attempt is rewarded, $R_1 = 0$. Furthermore, if we assume that all attempts are on the same site, the likelihood ratio grows following the recursive equation:

$$R_{t+1} = \frac{R_t + p_{SW}}{(1 - p_{SW})(1 - p_{RWD})} > R_t \quad (S14)$$

We can define the auxiliary quantity

$$\rho = \frac{1}{(1 - p_{SW})(1 - p_{RWD})} > 1. \quad (S15)$$

Our recursive equation becomes:

$$R_{t+1} = \rho \cdot (R_t + p_{SW}) \quad (S16)$$

This is a standard linear recursion that we can solve with linear transformation

$$S_t = R_t + \frac{p_{SW}}{\rho - 1} \quad (S17)$$

The recursion of S_t is:

$$S_t = \frac{p_{SW}}{\rho - 1} \quad (S18)$$

$$S_{t+1} = \rho \cdot S_t$$

whose solution is

$$S_t = p_{SW} \frac{\rho^{t-1}}{\rho - 1}, \quad (S19)$$

therefore:

$$R_t = p_{SW} \frac{\rho^{t-1} - 1}{\rho - 1} \quad (S20)$$

That is to say R_t grows exponentially with rate $\log(\rho) = -\log(1 - p_{RWD}) - \log(1 - p_{SW})$: increasing either p_{RWD} or p_{SW} would increase the growth rate of R .

DATA AND CODE AVAILABILITY

All analysis was performed using custom code written in Julia (Bezanson et al., 2017). The code used to simulate value or inference models is available on GitHub, under the MIT license, at <https://github.com/pieverb/ValueInferenceTools.jl>. The data is published on Zenodo with DOI [10.5281/zenodo.3607558](https://doi.org/10.5281/zenodo.3607558) and can be found at <https://zenodo.org/record/3607558>.

Supplementaries

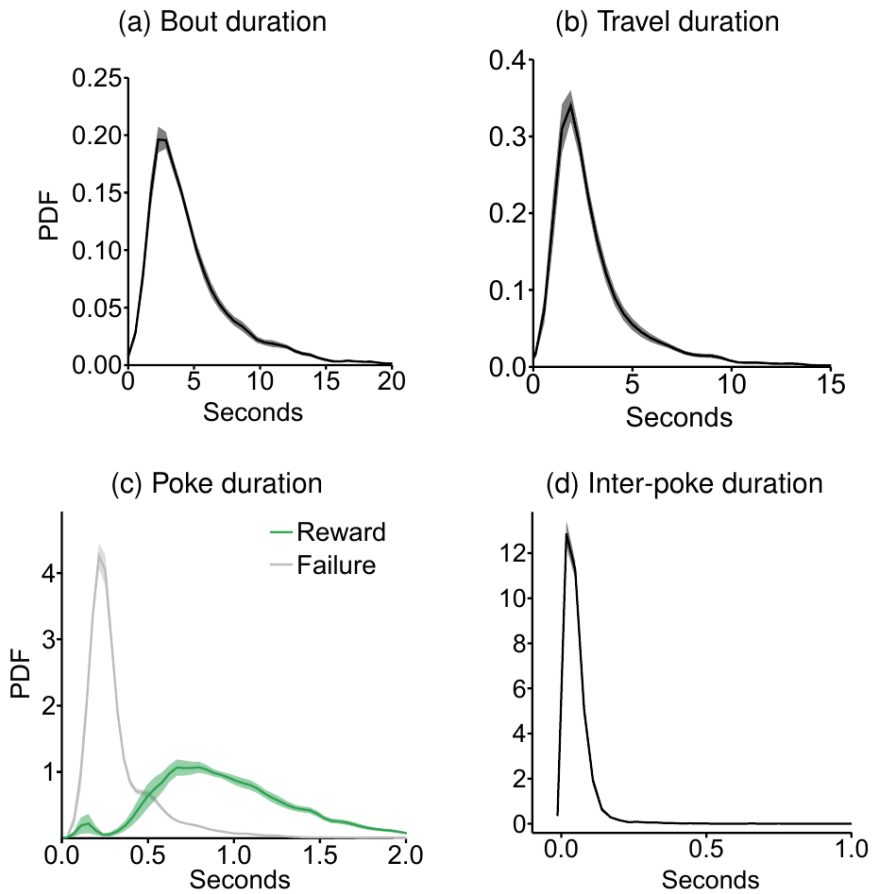


Figure S1. Related to Figure 2. Summary plots of behavioral times for trained animals (day 10 to 12). **a** Duration of poke bouts (from first poke-in to last poke-out before leaving the port). **b** Duration of travel (i.e. from last poke-out to first poke-in of next trial on the other site). **c** Time spent inside the poke for rewarded and unrewarded attempts. **d** Inter-poke interval within pokes on the same site (outliers of duration greater than 3 seconds have been excluded from the analysis for numerical reasons).

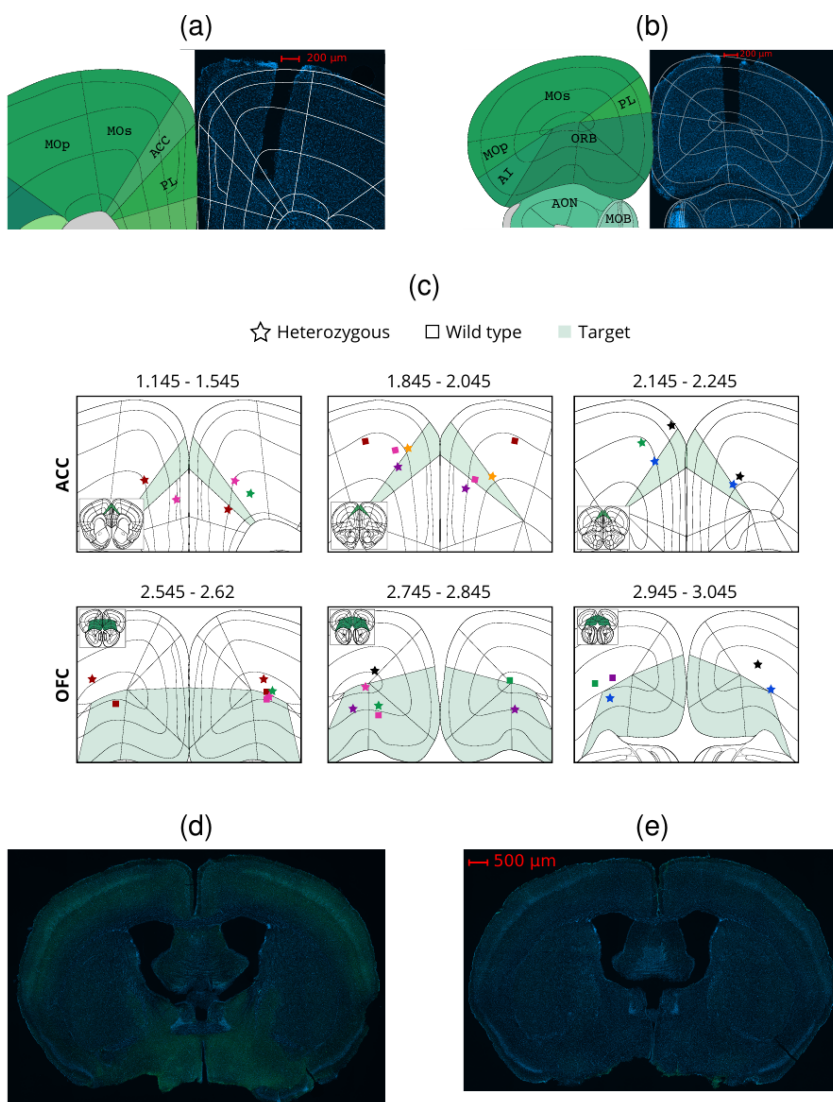


Figure S2. Related to Figure 5. Inhibition of ACC or OFC cortex in VGAT-ChR2 mice. **a, b** Coronal section from 2 VGAT-ChR2 mouse (blue, DAPI). Dark area show the location of the optic cannula over ACC in (a) or over OFC in (b). Correct positioning of the fiber was verified overlapping reference images from the coronal Allen Mouse Brain Reference Atlas. ACC, anterior cingulate; AI, agranular insular area; AON, anterior olfactory nucleus; CLA, claustrum; DP, dorsal peduncular area; fa, corpus callosum; ILA, infralimbic area; MOp, primary motor area; MOs, secondary motor area; PL, prelimbic cortex; ORB, orbital area; TT, taenia tecta. Image credit: Allen Institute. **c** Fiber tip placement in 9 out of 10 animals (in one wild type animal it was impossible to perform histological controls, see Table S1 for details) implanted in ACC (top row) and in the 10 animals implanted in OFC (bottom row). **d, e** Fluorescence widefield microscopy of EYFP signal conjugated to ChR2 in mouse VGAT-ChR2-EYFP line 8, in an example heterozygote (**d**) and an animal initially genotyped as heterozygous that was excluded from the dataset due to lack of EYFP expression (**e**).

MOUSE ID	SIDE	TARGET	AP	ML	DV	GENOTYPE	EYFP EXPRESSION
B21	Right	N/A	N/A	N/A	N/A	Wild type	N/A
B21	Left	N/A	N/A	N/A	N/A	Wild type	N/A
B22	Right	ACC	2.045	1.17	1.17	Wild type	FALSE
B22	Left	ACC	2.045	-1.17	1.18	Wild type	FALSE
B23	Right	ACC	2.145	0.83	1.99	Heterozygous	TRUE
B23	Left	ACC	2.245	-0.23	1.19	Heterozygous	TRUE
B24	Right	ACC	1.145	0.99	1.88	Heterozygous	TRUE
B24	Left	ACC	2.245	-0.69	1.45	Heterozygous	TRUE
B25	Right	ACC	1.42	0.73	1.66	Heterozygous	TRUE
B25	Left	ACC	1.42	-0.2	1.96	Heterozygous	TRUE
B41	Right	ACC	1.545	0.62	2.12	Heterozygous	FALSE
B41	Left	ACC	1.545	-0.71	1.65	Heterozygous	FALSE
B42	Right	ACC	2.145	0.72	2.12	Heterozygous	TRUE
B42	Left	ACC	2.145	-0.48	1.75	Heterozygous	TRUE
B43	Right	ACC	1.845	0.82	1.75	Heterozygous	TRUE
B43	Left	ACC	1.845	-0.5	1.3	Heterozygous	TRUE
B44	Right	ACC	1.945	0.38	1.93	Heterozygous	TRUE
B44	Left	ACC	1.945	-0.66	1.6	Heterozygous	TRUE
B45	Right	ACC	2.045	0.53	1.79	Wild type	FALSE
B45	Left	ACC	2.045	-0.69	1.32	Wild type	FALSE
P21	Right	OFC	2.62	1.26	2.13	Wild type	FALSE
P21	Left	OFC	2.745	-1.01	2.6	Wild type	FALSE
P22	Right	OFC	2.62	1.35	1.99	Heterozygous	TRUE
P22	Left	OFC	2.745	-1.01	2.44	Heterozygous	TRUE
P23	Right	OFC	2.745	1.22	2.5	Heterozygous	TRUE
P23	Left	OFC	2.745	-1.42	2.49	Heterozygous	TRUE
P24	Right	OFC	2.62	1.26	2.1	Wild type	FALSE
P24	Left	OFC	2.62	-1.19	2.2	Wild type	FALSE
P25	Right	OFC	2.62	1.3	2.08	Heterozygous	TRUE
P25	Left	OFC	2.845	-1.22	2.14	Heterozygous	TRUE
P41	Right	OFC	3.045	1.35	2.03	Heterozygous	TRUE
P41	Left	OFC	3.045	-1.22	2.17	Heterozygous	TRUE
P42	Right	N/A	N/A	N/A	N/A	Wild type	FALSE
P42	Left	OFC	3.045	1.18	1.85	Wild type	FALSE
P43	Right	OFC	2.545	-1.21	1.8	Heterozygous	TRUE
P43	Left	OFC	2.545	1.57	1.8	Heterozygous	TRUE
P44	Right	OFC	2.8455	-1.14	2.04	Wild type	FALSE
P44	Left	OFC	2.945	1.46	1.94	Wild type	FALSE
P45	Right	OFC	3.045	-1.14	1.63	Heterozygous	TRUE
P45	Left	OFC	2.845	1.07	1.88	Heterozygous	TRUE

Table S1. Related to Figure 5 and Figure S2. Optic fibers placement coordinates and expression EYFP.

Location of optic fibers across all VGAT-ChR2 animals used for this paper with anterior-posterior (AP), medio-lateral (ML) and dorso-ventral (DV) coordinates, according to the Allen Mouse Brain Atlas (Lein et al., 2007). In mouse B21, it was not possible to perform histological controls, due to the sudden death of the animal after the experiment period, which precluded us from performing the perfusion of its brain. In mouse P42, it was not possible to determine the placement of the fiber in the right hemisphere, due to damage in the slices during cutting. The expression of EYFP conjugated to ChR2 was assessed through widefield fluorescence microscopy to confirm the animal genotype. In mouse B41, EYFP signal was not detected (Fig. S2) despite the fact that the animal was initially genotyped as heterozygous; we therefore excluded it from the analysis.

Chapter 5: Maturation of prefrontal input to dorsal raphe increases behavioral persistence in mice

Nicolas Gutierrez-Castellanos, Dario Sarra, Beatriz Godinho & Zachary F. Mainen

Summary

The ability to persist towards a desired objective is a fundamental aspect of behavioral control and crucial for goal directed behaviors. In this study we describe the postnatal development of a long-range circuit underlying the emergence of behavioral persistence in mice. We found that top-down projections from the prefrontal cortex to the dorsal raphe nucleus mature between the third and the eighth postnatal weeks and this change correlates with the emergence of a more persistent behavioral phenotype in a self-paced foraging task, suggesting a causal link between this pathway and behavioral persistence. To further test this idea, we selectively ablated the prefrontal cortex neurons projecting to the dorsal raphe nucleus in adult mice and show that this procedure reverts them to a foraging behavioral phenotype similar to the juvenile animals. These results reveal the developmental dynamics of the anatomical substrate underlying the emergence of a fundamental cognitive ability, whose impairment is characteristic of several psychiatric disorders.

Introduction

The emergence of behavioral control, including attention, patience, cognitive flexibility, and behavioral persistence, occurs during critical periods of postnatal development. In these phases, environment and experience contribute to the maturation of higher cognitive functions (Mischel et al., 1989; Larsen and Luna, 2018; Tooley et al., 2021), which sets the foundations of future social and cognitive abilities during adulthood (Casey et al., 2011; Moffitt et al., 2011).

Ethologically, the development of behavioral control is critical for selective fitness and, thus, survival. For instance, in the natural environment, food resources are often sparsely distributed and depleted with consumption. Therefore, the well-known tradeoff between exploiting a depleting resource and exploring in search of alternatives is crucial to reach an optimal foraging strategy and obtain the maximum amount of resources with minimal waste of physical effort. Therefore, a forager in a possibly depleted patch of food faces an important dilemma - to stay or to leave - that calls for a careful balancing persistence and flexibility

(Charnov, 1976; Lottem et al., 2018; Morris and Davidson, 2000; Vertechi et al., 2020).

From a neural perspective, cognitive development correlates with large-scale synaptic and structural changes (e.g. Durston and Casey, 2006; Zuo et al., 2010, 2017) that are considered to underlie the emergence of increasing cognitive control over innate impulsive behavioral tendencies (Luna et al., 2001; Fair et al., 2009; Alexander-Bloch, 2013). A variety of evidence links the medial prefrontal cortex (mPFC) to the expression of behavioral control in a wide range of mammal species. For instance, humans and macaques with prefrontal cortical damage display deficits in behavioral flexibility, decision making, and emotional processing (Roberts et al., 1998; Rudebeck et al., 2013; Izquierdo et al., 2017), as well as a notable increase in impulsive behavior (Hitami and Uno, 2002; Berlin, 2004; Fellows, 2006; Dalley and Robbins, 2017). In line with this, local pharmacological inhibition of mPFC significantly limits rats' ability to wait for a delayed reward (Narayanan et al., 2006; Murakami et al., 2017).

Crucially, the mPFC undergoes intense postnatal maturation from childhood to adulthood, particularly during adolescence (Chini & Hanganu-Opatz., 2021), which in humans spans from years ~10-18 of life and in mice from weeks ~3-8 of life, and is a period of intense somatic maturation, including sexual development (Bell, 2018).

The maturation of the mPFC includes structural and functional modifications that overlap in time with the development of patience and behavioral inhibition during childhood and adolescence (Chini and Hanganu-Opatz, 2021; Sakurai and Gamo, 2019). Although it has been long hypothesized that the neural changes occurring in the mPFC during development are central to the emergence of behavioral control (e.g. Sowell et al., 1999; Durston and Casey, 2006), the specific plastic arrangements underlying behavioral control development remain poorly understood.

A number of studies have focused on the local changes of mPFC circuits, such as the changes in cortical thickness caused by cellular structural plasticity and synaptic pruning that are characteristic of early postnatal developmental phases in both humans and mice (e.g. Nagy et al., 2004; Alexander-Bloch, 2013; Ueda et al., 2015; Kolk & Rakic 2021),

as a putative locus underlying cognitive development. More recently, studies in rodents have shed light on the development of long-range top-down mPFC extracortical connections as the putative origin of certain aspects of cognitive development (Klune et al., 2021). In particular, the development of mPFC afferents to the amygdala may shape the response to threats across different stages of development (Arruda-Carvalho et al., 2017; Dincheva et al., 2015; Gee et al., 2016), and the development of mPFC input onto the dorsal raphe nucleus (DRN) shapes the response to stress (Soiza-Reilly et al., 2019).

The DRN is a brain region whose activity causally affects the animal's ability to wait and persist (Fonseca et al., 2015; Lottem et al., 2018; Miyazaki et al., 2012, 2014, 2018). A growing body of evidence supports that increasing 5HT neuron activity in the DRN enhances behavioral persistence not only in reward-seeking behaviors (Miyazaki et al., 2012, Fonseca et al., 2015, Lottem et al., 2018) but also in overcoming adverse situations (Nishitani et al., 2019; Ohmura et al., 2020).

The mPFC sends a dense glutamatergic projection to the DRN at the adult stage (Pollak Dorocic et al., 2014; Weissbourd et al., 2014; Zhou et al., 2017), which can bidirectionally modulate the activity of 5HT DRN neurons either through monosynaptic excitation or by disinaptic feedforward inhibition through local interneurons (Challis et al., 2014; Geddes et al., 2016; Maier, 2015; Warden et al., 2012). In particular Warden and colleagues have shown that selective optogenetic activation of the mPFC inputs to the DRN elicits active behavioral responses in a challenging context (Warden et al., 2012), and perturbations in the development of this pathway lead to maladaptive anxiety levels (Soiza-Reilly et al., 2019).

Given the reciprocal connectivity between the mPFC and DRN (Puig and Gullledge, 2011) and that both areas causally modulate animals' ability to wait for delayed rewards (Schweighofer et al., 2008; Miyazaki et al., 2011, 2014, 2018; Fonseca et al., 2015; Murakami et al., 2017; Ciaramelli et al., 2021), it seems plausible that the maturation of mPFC input to the DRN over development could underlie the emergence of behavioral persistence in mice.

Therefore, we sought to characterize the development of cortical innervation onto the DRN and its functional consequences in the context of behavioral control. In contrast to previous studies in which adult behavioral readouts were assessed after developmental perturbations (Bitzenhofer et al., 2021; Soiza-Reilly et al., 2019), we undertook a longitudinal study, characterizing behavior, synaptic physiology and anatomy, in parallel, from adolescence to adulthood.

The study of cognitive development in mice is challenged by the fact that most tasks designed to assess cognition require several weeks or even months of training (Pittaras et al., 2020; Sanchez-Roige et al., 2012; The International Brain Laboratory et al., 2021; Winstanley and Floresco, 2016). Here, we took advantage of the innate foraging abilities to test naive mice in a two-port probabilistic foraging task which measured their persistence in exploiting a given foraging port before exploring the other. This task was used previously in adult mice to test their ability to infer the hidden state of the foraging site, a behavior that took several days to establish (Vertechi et al., 2020). Here, instead, we made use of the fact that even naive mice performed the task competently, although sub-optimally, from the first day of exposure. We found that all mice were able to perform the movements required to obtain water within minutes of entering the behavioral box, including nose-poking in the ports, licking at the water spout, and alternating between ports, suggesting that this task likely taps into a largely innate behavioral repertoire.

Our results showed that juvenile mice (3-4 weeks old) are less persistent than adult mice, that these behavioral changes match the development of the mPFC to DRN projection, and that ablation of this projection in adult mice recapitulates the juvenile behavioral phenotype. These findings suggest that the development of top-down PFC-DRN afferents is critical to the emergence of cognitive control over behavioral impulsivity that characterizes adulthood.

Results

Cortical top-down input over the dorsal raphe matures in the transition between adolescence and adulthood in mice.

First, to characterize the development of neocortical projections to the DRN, we focused on the afferents of layer V neurons, which are the primary origin of these projections (Pollak Dorocic, 2014). Using a mouse line expressing channelrhodopsin-2 (ChR2) in a large fraction neocortical layer V neurons (Rbp4-Cre/ChR2-loxP) (Leone et al., 2015), we performed ChR2-assisted circuit mapping (sCRACM) (Petreanu et al., 2009) of cortical afferents in brain slices containing the DRN obtained from mice between postnatal weeks 3 to 12 (Fig 1A). Taking advantage of the fact that ChR2-expressing axons are excitable even when excised from their parent somata, we evoked firing of presynaptic ChR2-expressing cortical axons innervating the DRN while recording the electrophysiological responses of postsynaptic DRN neurons. We assessed the fraction of recorded DRN neurons receiving cortical excitatory synaptic input (connection probability, Pcon) and the strength of this connection (amplitude of the evoked synaptic response) at different developmental time points.

We found a dramatic increase in the connection probability and amplitude of cortico-raphé input between weeks 3 and 8 (Fig. 1B-C). Between the 3-4 weeks (juvenile mice), the probability of DRN neurons receiving cortical input was equal to 0.11. This probability increased significantly to 0.66 (Pcon 3-4 weeks vs. Pcon 5-6 weeks, Chi-Square test χ^2 (1, N = 53 neurons) = 24.09, $p = 0.00001$) between weeks 5 and 6, reaching a peak connection probability of 0.82 between weeks 7 and 8 (Fig. 2C). Between 5-6 and 7-8 weeks (i.e. late juvenile to adult mice), the amplitude of the optogenetically evoked currents increased from 26 ± 5.9 pA to 123 ± 14.5 pA (mean $\pm$ SEM, two-tailed t-test, $t(57) = 4.028$, $p = 0.002$), suggesting a gradual maturation of the cortico-raphé pathway up to week 8. To test whether there is a further development of this pathway in the later stages of development, we recorded slices from 12-week old mice. We observed no further increase in either the connection probability (Pcon 7-8 weeks = 0.82 vs. Pcon 5-6 months = 0.80, Chi-Square test χ^2 (1, N = 70 neurons) = 0.03, $p = 0.84$) or the

input magnitude (7-8 weeks old = 123 ± 14.5 pA vs. 5-6 months old = 110 ± 8.7 pA, two-tailed t-test, $t(60) = 1.27$, $p = 0.21$, Fig. 2B-C). Altogether, these results suggest a gradual maturation of the cortico-raphe pathway between weeks 3 and 8, that plateaus afterwards. Importantly, the location of the recorded DRN neurons was comparable between juvenile and adult mice (Fig. S1) and thus, the connectivity changes observed across development do not reflect a biased sampling of differentially innervated sub-regions of the DRN. Furthermore, we observed a comparable membrane resistance (3-4 weeks: median = 400.4, 95% CI = [243.3, 500], 5-6 weeks: median = 427.4, 95% CI = [294.1, 925.2], 7-8 weeks: median = 485.4, 95% CI = [409.8, 699.3], 5-6 months: median = 546.4, 95% CI = [348.4, 806.5], Kruskal-Wallis $H(3) = 3.58$, $p = 0.17$) and membrane capacitance (3-4 weeks: median = 23.8, 95% CI = [15, 29.5], 5-6 weeks: median = 22.2, 95% CI = [12.7, 26.4], 7-8 weeks: median = 27.5, 95% CI = [19.3, 34.5], 5-6 months: median = 26.2, 95% CI = [11.8, 40.1], Kruskal-Wallis $H(3) = 2.01$, $p = 0.37$) in DRN neurons over development (Fig. S2A,B), suggesting that changes at the level of the passive propagation of current through DRN neurons is not the underlying cause of the increase in connection probability and input magnitude observed over time.

In these experiments, the onset of ChR2 expression is dictated by the Cre recombinase expression under the control of the native Rbp4 promoter over development. Therefore, if in the juvenile cortex there were fewer neurons expressing Rbp4 or the onset of expression was near our recording time point, this could affect the net amount of ChR2-expressing top-down cortical axons and/or their net excitability. To control that our findings reflect a development process and not a genetic artifact caused by the temporal dynamics of Rbp4 expression, we performed two additional control experiments in one of the main cortical origins of afferents onto the DRN, the mPFC (Weissbourd et al., 2014, Zhou et al., 2017).

First, we compared the density of neurons in the mPFC expressing a fluorescent reporter (tdTomato) under the control of the Rbp4 promoter in juvenile and adult mice. The same density of tdTomato expressing somas was detected in the mPFC of juvenile and adult Rbp4-Cre\tdTomato-loxP mice (Juveniles: median = 4.79, 95% CI = [3.03, 7.34], vs. Adults: median = 4.79, 95% CI = [3.03, 7.34], Mann

Whitney U test ($N_{\text{Juveniles}} = 3, N_{\text{Adults}} = 4$) = 6, $p = 0.99$, Fig. S2C), indicating that a comparable number of neurons underwent a Cre-dependent recombination of the tdTomato fluorescent reporter under the control of the Rbp4 promoter at both developmental time points. Second, we compared the light-evoked somatic current produced in layer V neurons expressing ChR2 under the Rbp4 promoter in juvenile and adult mice. In agreement with the previous control, layer V neurons in the mPFC expressing ChR2 under the Rbp4 promoter produced the same amount of photocurrent upon light stimulation in juvenile and adult mice (Repeated Measures ANOVA, $F(1, 11) = 0.1379$, $p = 0.71$ for age factor, Fig. S2D). These experiments show that juvenile and adult mice have similar densities of cortical layer V projection neurons that could give rise to DRN afferents and that these neurons express similar amounts of ChR2 and thus, if present, projections should be equally detectable by optogenetic circuit mapping across ages. Altogether, this evidence suggests that the maturation of cortico-raphé projections we report is caused by a developmental process and is not explained by experimental artifacts.

To further understand the mechanisms underlying the cortico-raphé input strengthening observed over development, we investigated whether the changes of connection probability and input amplitude we observed were accompanied by differences in the density of cortical axonal innervation over the DRN. Indeed, we observed a significantly higher density of Rbp4 positive axons around the DRN in adults compared to juveniles (2.2 ± 0.47 vs. 5.7 ± 0.56 axons per bin in juvenile vs. adult mice, two-tailed t-test, $t(9) = 4.15$, $p = 0.002$, Fig. 1D-E). This observation supports the idea that the increase in physiological strength we observed reflects in part the growth of new connections between the neocortex and DRN.

To assess whether the development of cortico-raphé projections is specific to raphe projecting cortical afferents or it reflects a more general maturation of corticofugal projections over adolescence in mice, we mapped the anatomical and synaptic development of cortico-accumbens projections, which are mainly originated in the PFC (Phillipson & Griffiths 1985, Li et al., 2018) and which functional connectivity has been previously assessed in juvenile rodents (Gorelova and Yang, 1996). In contrast to cortico-raphé afferents,

cortico-accumbens projections did not undergo any significant structural change over the same developmental period (4.5 ± 0.54 vs. 5.8 ± 0.74 axons per bin in juvenile vs. adult mice, two-tailed t-test, $t(8) = 1.09$, $p = 0.40$, Fig. 1D-E). Consistent with the anatomy, the Chr2-assisted mapping of cortico-accumbens connections in juvenile and adult Rbp4-Chr2 mice revealed no change in either the connection probability (Pcon 3-4 weeks = 0.90 vs. Pcon 7-8 weeks = 0.87, Chi-Square test $\chi^2(1, N = 19 \text{ neurons}) = 0.03$, $p = 0.81$) or the input amplitude in the transition from juveniles to adults (two-tailed t-test, $t(15) = 0.15$, $p = 0.88$, Fig. 1F). Altogether, these observations reveal the structural and synaptic development of a sub-population of cortical afferents targeting the DRN during the transition from juvenility to adulthood in mice that does not reflect a generalized development of corticofugal projections.

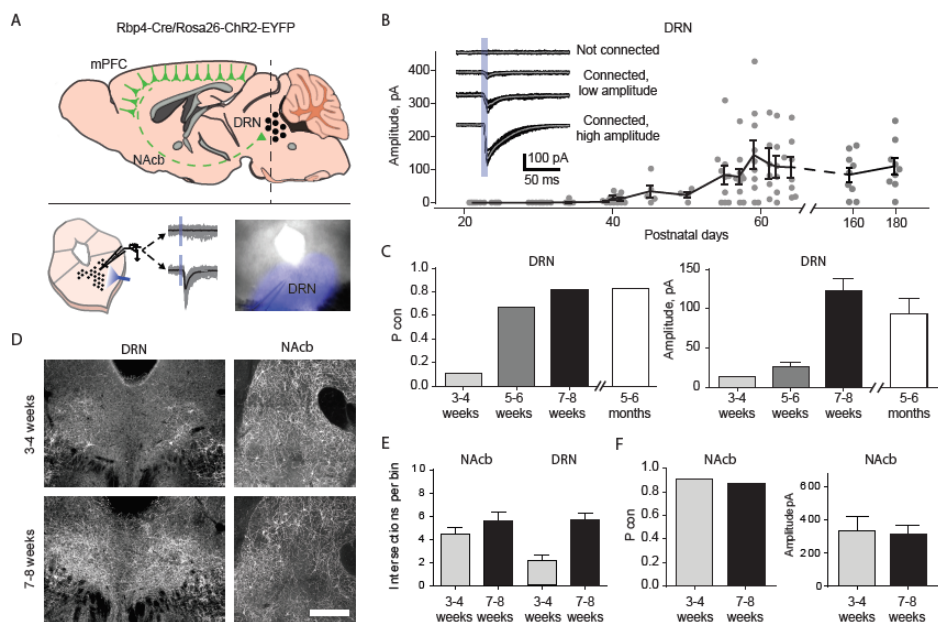


Figure 1. Top-down cortico-raphe connections develop over adolescence in mice. (A) Schematic representation of a sagittal view of an Rbp4-Chr2 mouse brain illustrating top-down cortico-raphe afferents. Coronal slices containing the DRN were obtained ex vivo, and whole-cell recordings of DRN neurons were performed to assess cortical connectivity upon light stimulation. (B) Optogenetically evoked EPSCs were recorded in DRN neurons contacted by Chr2 expressing cortical axons (122 neurons, 20 Rbp4-Chr2 mice). The current amplitude of cortico-raphe connections is plotted as a function of postnatal age in mice (C) Pooled connection probability (Connected cells/ total cells) and averaged connection amplitude of

cortico-DRN afferents at four different developmental points: Early juvenile (3-4 weeks), late juvenile (5-6 weeks), Early adult (7-8 weeks), late adult (5-6 months). (D) Example images illustrate an increased cortico-DRN innervation in adult mice compared to juveniles, while the cortico-accumbens innervation remains constant over the same time period. Scale bar = 400 microns. (E) Number of axonal intersections quantified in the DRN and nucleus accumbens of juvenile and adult mice. (F) Pooled connection probability and averaged connection amplitude of cortico-accumbens afferents in early juvenile and early adult mice.

Baseline persistence correlates with the maturation of cortico-raphe input in the transition between adolescence and adulthood in mice.

To investigate the development of behavioral persistence in mice, we employed a previously published self-paced foraging task (Vertechi et al., 2020). The setup consists of a box with two nose-ports separated by a barrier (Fig. 2A). Each nose-port constitutes a foraging site that mice can actively probe in order to receive water rewards. Only one foraging site is active at a time, delivering reward with a certain probability ($P_{\text{rwd}} = 90\%$) when probed. Each try in the active site can also cause a switch of the active site with a certain probability ($P_{\text{sw}} = 90\%$) (Fig. 2B). After a state switch, mice have to travel to the other port to obtain more reward, bearing a time cost to travel. While this task was previously studied after multiple days of training, after which adult mice use an inference-based strategy, early in learning, they show a value-based strategy, staying longer when they receive more rewards (Vertechi et al., 2020). This behavior was useful for the purpose of assessing cognitive development, as it requires persistence in poking at the port despite reward failures.

To measure how developmental changes affect mice's behavioral persistence, we compared the behavior of juveniles (weeks 3-4) and adult mice (weeks 7-8) on their first exposure to the apparatus and task. Presumably due to the novelty of the apparatus, mice tended to interleave poking in the port with investigating the surrounding, often taking long pauses in between pokes at the same port, resulting in a less regular poking pace than experienced mice (Fig. S3 A-C). It is unlikely that the animals are tracking the number of pokes to decide when to visit the other side. Therefore to assess mice's behavioral persistence, we

assessed the animal *leaving time*, measured as the overall time spent investigating one port before visiting the other (Fig. 2B). First, we assessed whether the poking behavior of juveniles and adults presented some gross qualitative alteration. To assess this we measured the proportion of time spent poking during each trial and we found no difference between the two groups (Fig 2C; Adults: median = 0.48, 95% CI = [0.41, 0.55], Juveniles: median = 0.45, 95% CI = [0.38, 0.54]; Mann Whitney U test ($N_{\text{Adults}} = 21$, $N_{\text{Juveniles}} = 23$) = 268, $p = 0.54$).

Next, we compared how likely to leave juveniles and adults are as a function of the time elapsed from the beginning of a trial (Fig. 2D). We observed a leftward shift in the leaving probability sigmoidal function of the juveniles compared to the adults, indicative of reduced behavioral persistence. To test this hypothesis, we performed a logistic regression to explain the decision to leave after each poke and identify which factors significantly contribute to more or less behavioral persistence. The individual variability was accounted for through generalized linear mixed models with random intercept and slopes for each mouse (see methods). A factor was considered to significantly affect the decision to leave if the value of its estimated coefficient plus 95% confidence interval (1000 parametric bootstrap analysis, see methods) did not cross 0. Results indicate three significant main factors (Fig. 2E). First, as expected, the elapsed time from the beginning of the trial (PokeTime) significantly increased the probability of leaving the patch. Second, the elapsed number of trials from the beginning of the session (Trial) also increases the probability of leaving, indicating that animals become less persistent over the course of the session (Fig. S3D). This effect could suggest a drop in motivation due to the water drunk during the session. Third, we found a significant effect of the categorical age factor. Using adults as a baseline condition, this analysis showed that juveniles are significantly more likely to leave the current site to explore the other, indicating less behavioral persistence. Interactions between age and the other main factors were also allowed, and we only found a significant negative interaction between juveniles and PokeTime, indicating that the probability of leaving as a function of time elapsed in a trial increases more slowly for juveniles than adults. We verified that the model ability to explain the decision to leave is significantly improved by considering the Age factor (likelihood ratio test on $\text{Leave} \sim 1 + \text{PokeTime} + \text{Trial} + \text{Age} + \text{PokeTime} \& \text{Age} + \text{Trial} \& \text{Age} + (1 + \text{PokeTi}$

me+Trial|MouseID) versus *Leave~1+PokeTime+Trial+(1+PokeTime+Trial|MouseID)*: $X^2_{(3)} = 19.27$, $p = 2e-4$).

We next examined whether the juvenile phenotype resulted in an increase in the number of incorrect trials, defined as leaving before the rewarded port switches state. Indeed compared to adults, juveniles performed a larger proportion of incorrect trials (Fig. 2F; Adults: median = 0.24, 95% CI = [0.19, 0.31], Juveniles: median = 0.35, 95% CI = [0.28, 0.55]; Mann Whitney U test($N_{\text{Adults}} = 21$, $N_{\text{Juveniles}} = 23$) = 137.5, $p = 0.015$, effect size = 0.72), supporting the idea that the development of higher behavioral control leads to a reduction of impulsive actions.

Finally, we investigated whether developmental differences are consistent in male and female mice. We found no significant effect on the probability of leaving of the sex factor either alone or in interaction with animals' age (Fig. S3E), and including sex as a predictor had no significant improvement in the model's ability to explain the decision to leave (Fig. S3F, likelihood ratio test on *Leave~1+PokeTime+Trial+Age+Sex+Age&Sex+(1+PokeTime+Trial|MouseID)* versus *Leave~1+PokeTime+Trial+Age+(1+PokeTime+Trial|MouseID)*: $X^2_{(2)} = 3.46$, $p = 0.18$). These results suggest that the development of persistence occurs at a similar rate in male and female mice.

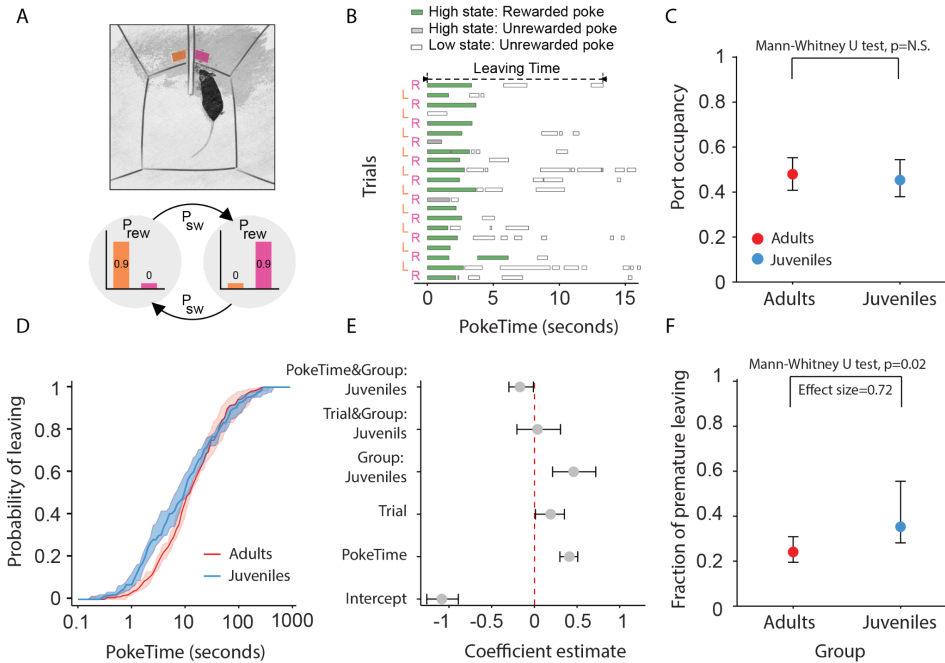


Figure 2. Adult mice persist longer than juveniles in exploiting a foraging patch (A) Illustration of the rodent foraging task. Water-deprived mice seek rewards by probing two nose-ports. (B) An example of poking behavior over time. Pokes in the active state can be rewarded (in green) or not (in gray). Pokes in the inactive state are never rewarded (in white). After the state switches, the mice have to travel to the other side (left or right port, L and R) to obtain more water. (C) Fraction of time spent in the nose port per trial (median $\pm$ 95% CI across mice) for adult mice (blue, $n = 21$) and juveniles (orange $n = 23$). (D) Cumulative distribution of the probability of leaving (median $\pm$ 95% CI across mice) as time elapses from the first poke in a trial for adults and juvenile animals. (E) Regression coefficients $\pm$ 95% CI resulting from a parametric bootstrap ($n = 1000$) of a mixed models logistic regression to explain the probability of leaving. Among other factors, the age group significantly affects the decision to leave, with juveniles more likely to leave than adults, indicative of reduced behavioral persistence. (F) Fraction of incorrect trials (median $\pm$ 95% CI across mice), defined as leaving the port prematurely before the state switches.

mPFC-DRN pathway ablation in adult mice recapitulates juvenile behavioral features.

The above results establish a correlation between the development of the descending cortical input to the DRN and the

emergence of behavioral persistence. To more directly causally link the development of cortico-raphé afferents to the increase in persistence observed in the probabilistic foraging task, we next ablated the cortico-raphé pathway in adult mice and assessed the impact on behavioral persistence.

To ablate cortico-raphé afferents, we used an engineered version of Caspase3 (taCasp3-TEVp) that is able to trigger apoptosis bypassing cellular regulation upon activation by the TEV protease, which is coexpressed in the same construct (Yang et al., 2013). We packaged a Cre dependent taCasp3-TEVp construct (or the reporter tdTomato as a control) in a retrogradely travelling AAV vector (rAAV), that we locally delivered in the DRN of Rbp4-Cre mice. This approach resulted in the fluorescent tagging of cortico-raphé layer V projecting neurons in control mice (tdTomato mice) and in the ablation of the same corticofugal pathway in taCasp3-TEVp injected mice (Caspase mice) (Fig. 3A).

The prelimbic/infralimbic (PL/IL) and anterior cingulate (AC) cortices, which constitute the mPFC, were the areas with the highest density of DRN-projecting tdTomato+ somas in control animals and consistently more extensive neuron density loss in caspase injected mice, quantified using the pan-neuronal marker NeuN (Fig. 3A-C, Fig. S4, control vs. caspase, two-sample Kolmogorov-Smirnoff Test = 0.026, $p = 0.005$ for PL/IL and $D = 0.022$, $p = 0.01$ for AC). We also found tdTomato+ somata in the medial orbitofrontal cortex (MO) of the control group; however, this projection was weaker in terms of tdTomato+ labelled neurons and, consistently, the difference in layer V NeuN densities between control and caspase mice was not significant (Figs. 3C, S4, $D = 0.016$, $p = 0.09$).

Apart from the mPFC, sparse labeling of tdTom+ neurons was found in more posterior levels of the neocortex, namely in the retrosplenial cortex (RS) and in the temporal association cortex (TeA) (Fig. S4). Nonetheless, tdTom+ neurons in the RS and TeA were found in only 5 and 3 out of 8 control animals, respectively. Consistently, the reduction in NeuN layer V neuronal density in these two areas was minimal and non-significant compared to controls (Fig. 3C, $D = 0.015$, $p = 0.16$ for RS and $D = 0.013$, $p = 0.14$ for TeA). In addition, no differences in NeuN density were observed between caspase injected

animals and controls in a control area not showing tdTomato expressing somas and therefore not projecting to the DRN (M1, Fig. 3C, $D = 0.008$, $p = 0.29$). These observations suggest that our ablation approach significantly affected mPFC-DRN projecting neurons, particularly from PL/IL and AC cortices.

When investigating the distribution of tdTomato expressing somas, we observed weak collateral projections of the cortical subpopulation projecting to the DRN in the lateral septum, lateral hypothalamic nucleus, the ventral tegmental area and the anterior periaqueductal gray; medium collateral axonal density in anterior subcortical olfactory nuclei (anterior dorsal endopiriform, anterior olfactory nucleus, dorsal taenia tecta and islands of Calleja) and the substantia nigra; and heavy collateralization in the dorsomedial striatum (Fig. S5).

We then assessed the impact of ablation of the mPFC-DRN projection on behavioral persistence using the foraging paradigm. We observed a similar pattern of differences between caspase and tdTomato mice as between juveniles and adults. Caspase and tdTomato animals did not differ in the proportion of time spent poking per trial (Fig. 3D; tdTomato: median = 0.45, 95% CI = [0.42, 0.51], Caspase: median = 0.55, 95% CI = [0.46, 0.78]; Mann Whitney U test ($N_{tdTomato} = 8$, $N_{Caspase} = 7$) = 13, $p = 0.09$). Caspase animals showed an increase in the probability of leaving the port, seen as a leftward shift in the cumulative distribution of leaving times (Fig. 3E). We applied the same logistic regression analysis to identify which factors significantly affect the probability of leaving. Similar to the previous result, both elapsed PokeTime and Trial significantly influenced the probability of leaving (Fig. 3F). Crucially, we found that animals lacking cortico-raphe projections are significantly more likely to leave the patch than control animals; however, this effect did not interact with PokeTime (Fig. 3F). Including the virus group factor significantly improved the explanatory power of the model (likelihood ratio test on $Leave \sim 1 + PokeTime + Trial + Virus + PokeTime \& Virus + Trial \& Virus + (1 + PokeTime + Trial | MouseID)$ versus $Leave \sim 1 + PokeTime + Trial + (1 + PokeTime + Trial | MouseID)$: $X^2_{(3)} = 10.84$, $p = 0.012$). Finally, we assessed whether the increased probability of leaving significantly increased the fraction of premature actions

undertaken by the caspase animals. Indeed compared to tdTomato mice, caspase mice committed a significantly larger proportion of errors, prematurely leaving the site before the active port switches (Fig. 3G; tdTomato: median = 0.22, 95% CI = [0.16, 0.33], caspase: median = 0.50, 95% CI = [0.32, 0.57]; Mann Whitney U test ($N_{\text{tdTomato}} = 8$, $N_{\text{Caspase}} = 7$) = 137.5, $p = 0.01$). Together these results indicate that ablating the mPFC-DRN pathway recapitulates the behavioral features characteristic of juvenile mice in the same foraging task, indicating that the mature mPFC-DRN projection is necessary for the behavioral persistence displayed by adult mice and that this pathway is likely to mediate at least in part the development of behavioral persistence in mice.

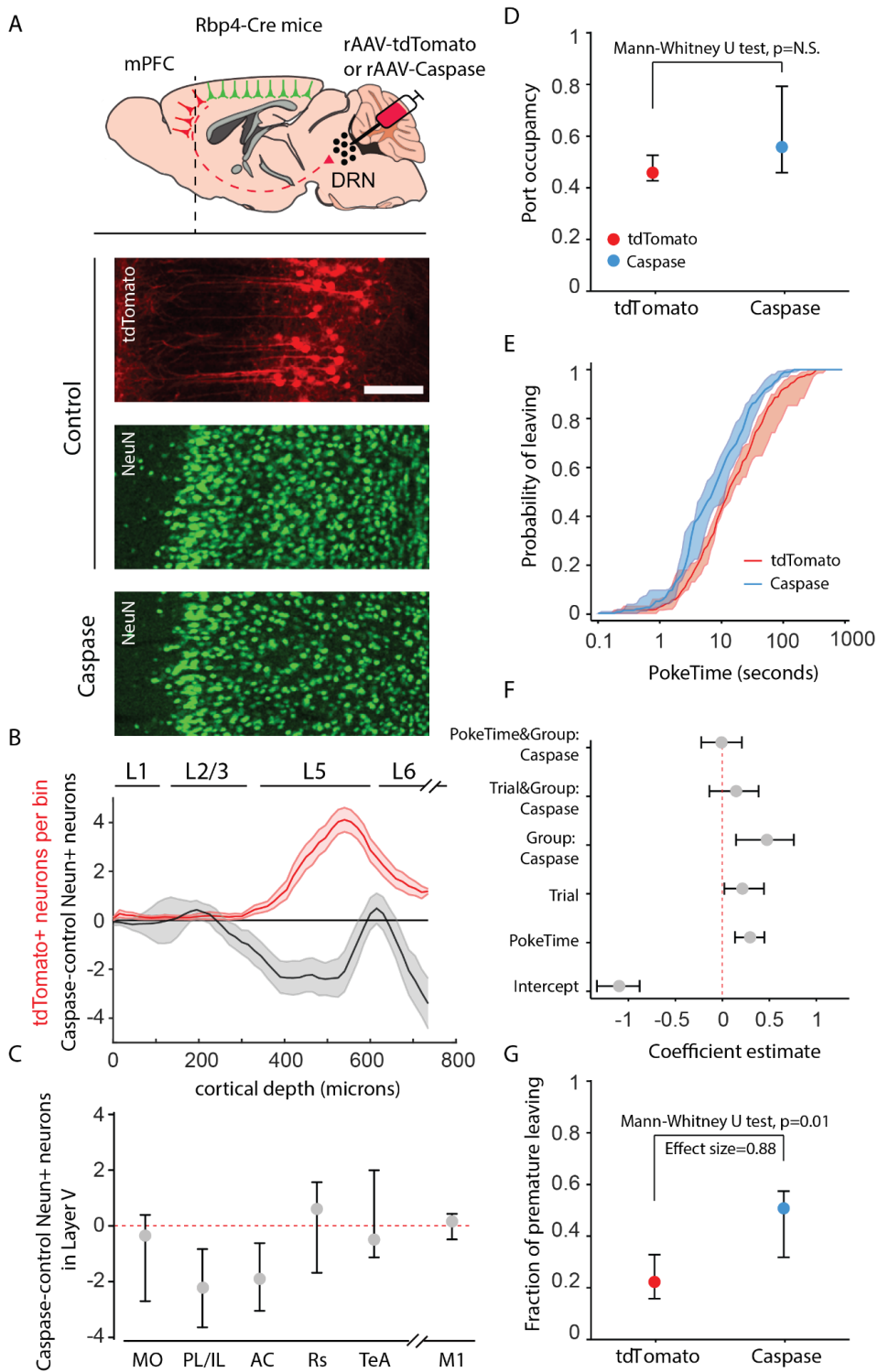


Figure 3. Animals lacking cortico-raphe projections show less behavioral persistence in exploiting a foraging patch.

(A) Schematic representation of the ablation strategy used for behavioral assessment. Retrogradely transporting AAV vectors expressing either the fluorescent reporter tdTomato (rAAV-tdTomato, tdTomato group, $n = 8$) or the intrinsically active apoptosis triggering Caspase3 (rAAV-Caspase, Caspase group, $n = 8$) were locally delivered in the DRN of Rbp4-Cre mice. In the cortical areas containing tdTomato expressing neurons in control animals (in example picture, PL/IL cortex) the density of neurons was quantified with an immunohistochemistry protocol against the pan-neuronal marker NeuN and compared to the neuronal densities obtained in the same cortical areas of ablated mice (scale bar = 200 microns). (B) Distribution of the neuronal density difference between ablated mice and the mean density of control mice per cortical depth bin in the PL/IL cortex (black shaded error plot). The neuronal density loss observed in ablated mice when compared to control NeuN densities matches the cortical depth in which tdTomato neurons are located (red shaded area). Shaded error plots represent mean $\pm$ SEM. (C) Summary of caspase-control NeuN density per brain area (MO: median = -0.36, 95% CI = [-2.696, 0.3988], PL/IL: median = -2.16, 95% CI = [-3.640, -0.8304], AC: median = -1.78, 95% CI = [-3.045, -0.6161], Rs: median = 0.67, 95% CI = [-1.802, 1.817], TeA: median = -0.15, 95% CI = [-1.002, 2.169], and M1: median = 0.04, 95% CI = [-0.6702, 0.2476]). (D) Median $\pm$ 95% CI fraction of time spent in the nose port per trial. The comparison between tdTomato (blue, $n = 8$) and Caspase (orange $n = 7$) showed no significant difference between the two groups (tdTomato: median = 0.45, 95% CI = [0.42, 0.51], Caspase: median = 0.55, 95% CI = [0.46, 0.78]; Mann Whitney U test ($N_{tdTomato} = 8$, $N_{Caspase} = 7$) = 13, $p = 0.09$). (E) Cumulative distribution of the probability of leaving as a function of trial time elapsed (median $\pm$ 95% CI across mice) for tdTomato and Caspase animals. (F) Regression coefficients $\pm$ 95% CI resulting from a parametric bootstrap ($n = 1000$) of a mixed models logistic regression to explain the probability of leaving. (G) Median $\pm$ 95% CI fraction of incorrect trials. Caspase animals do a significantly higher proportion of premature leaving.

Discussion

In the present study, we described how the postnatal maturation of the mPFC projection to the DRN during adolescence is linked to the performance of a probabilistic foraging task. Over the same period of development, the mPFC to DRN projection underwent a dramatic increase in potency and mice developed an increase in persistence in foraging behavior. Ablation of the mPFC-DRN pathway in adult mice recapitulated the features observed in the behavior of juvenile mice, supporting a causal relationship between the mPFC-DRN projection and behavioral persistence.

In a wide variety of species, including mice, adolescence corresponds to the emancipation from the parents (Spear, 2000), a

period in which individuals need to develop or refine skills to become independent. This ethological scenario may explain the evolutionary selection of juvenile behavioral traits (e.g. Sercombe, 2014), such as increased impulsivity or high risk taking behavior (Laviola et al., 2003). However, the abnormal development of cognitive control over the intrinsic behavioral tendencies of juveniles may underlie aspects of the etiopathology of impulsive and addictive disorders in adult humans (Wong et al., 2006; Reiter et al., 2016). In line with pre-adolescent humans' lack of delay gratification ability (Mischel et al., 1989), and with studies assessing impulsive behavior in mice over development (Sasamori et al., 2018), we found that mice of 3-4 weeks of age tend to be less persistent than 7-8 weeks old mice in a probabilistic foraging task. This led to a negative impact on foraging efficiency, with more premature site-leaving decisions..

From a neural perspective, maturational changes of prefrontal cortical areas, including the mPFC, have been previously linked to the emergence of cognitive skills during development in primates and humans (Nagy 2004; Velenova 2008; Luna et al., 2015). Such changes result in an increased top-down behavioral control and increased functional connectivity with cortical and subcortical targets (Hwang, 2010) in the transition between childhood to adulthood. However, the specific contribution of long-range top-down mPFC circuits and the cellular mechanisms underlying its development had not been previously investigated.

Here, using optogenetic-assisted circuit mapping we characterized the structural and functional development of cortico-raphé projections that take place over adolescence in mice. A recent report showed that a subpopulation of DRN-projecting mPFC neurons increases their axonal contacts over the DRN in an earlier phase of postnatal development (P2 to P14) (Soiza-Reilly et al., 2019). We found a mPFC-DRN connection strength at 5-6 weeks postnatal similar to that reported by Soiza-Reilly and colleagues at a similar developmental time point (4-5 weeks of age). In addition, we found a connection strength at 7-8 weeks of age similar to those reported in adult rodents elsewhere (Geddes et al. 2016; Zhou et al. 2017). Thus, our findings are consistent with previous observations in the literature and suggest that the maturation of mPFC-DRN afferents starts early in postnatal development

and undergoes an extended development period, plateauing only after reaching 7-8 weeks of age. Among the previous studies investigating the postnatal development of top-down afferents from the mPFC in rodents (Klune et al., 2021, Peixoto et al., 2016, Ferguson & Gao 2015), the latest mPFC afferent maturational process reported, the mPFC innervation over the basolateral amygdala, occurs up to week 4 (Arruda-Carvalho et al., 2017). Thus, to our knowledge, the mPFC-DRN pathway represents the latest top-down pathway from the cortex to develop.

Importantly, we found that the structural development of mPFC-DRN projections is causally linked to the maturation of behavioral persistence in adult mice. Using a genetically driven ablation approach (Yang et al., 2013), we selectively eliminated layer V cortical neurons projecting to DRN in adult mice. The procedure resulted in a behavioral phenotype that replicated key features of the juvenile foraging behavior. We observed a reduction in behavioral persistence, coupled with an increased fraction of errors. Furthermore, we localized the origin of these projections and quantified the local neuronal loss. The PL, IL, and AC cortices, areas that comprise the so called mPFC (Klune et al 2021), suffered a significant loss with the procedure, highlighting their importance for displaying behavioral persistence necessary for reward exploitation in a foraging task.

Previous reports have shown that the pharmacological inactivation of the IL cortex reduces response persistence in a foraging task (Verharen, 2020). Moreover, lesions of the IL cortex improves performance on a reversal learning task (Ashwell and Ito, 2014), which resembles the increased behavioral flexibility observed in juvenile mice (Johnson and Wilbrecht., 2011). In addition, neurons in the PL cortex have been shown to track reward value and reflect impulsive choices (Sackett et al., 2017). More generally, the inactivation of PL/IL cortices using optogenetics leads to an increase in premature responses in a probabilistic reversal task (Nakayama), while the optogenetic activation of the PL/IL cortices increases food-seeking behavior while reducing impulsive actions (Warthen, 2016).

Furthermore, lesions in the AC impair behavioral inhibition producing an increase in premature actions in rodents (Muir, Everitt, and

Robbins, 1996; Hvoslef-Eide and Bussey, 2018). More recently, it has been shown that the control of impulsive actions exerted by the AC requires intact signalling through Gi-protein in its layer-5 pyramidal neurons (Van der Veen et al., 2021). Altogether, there is considerable evidence linking the activity of the areas composing the mPFC (PL, IL and AC cortices) to the control of impulsive actions.

In addition, the optogenetic activation of DRN 5HT neurons, a major subcortical target of mPFC projections (Geddes et al., 2016; Zhou et al., 2017; Weissbourd et al., 2014, Pollak Dorocic et al., 2014), improves the performance of a delayed response task (Miyazaki et al., 2014, 2018; Fonseca et al., 2015) through an increase in active behavioral persistence (Lottem et al., 2018), which is the converse effect of the pharmacological silencing of the mPFC (Narayan and Laubach, 2006; Narayan et al., 2013; Murakami et al., 2017). Altogether, the emerging picture suggests that the individual activation of either mPFC or DRN converges into a behaviorally persistent phenotype. Consistently, the activation of mPFC-DRN top-down projections also has been shown to increase active persistence (Warden et al., 2012). However, previous studies have reported a net inhibitory effect of mPFC input onto 5-HT neurons in the DRN (Celada et al., 2001; Maier, 2015), particularly after prolonged trains of high-frequency stimulation (Srejic et al., 2015). This raises a question on the directionality with which mPFC input modulates DRN neurons activity in the context of behavioral control. One possible mechanism would be a frequency dependency on the net effect of mPFC input on the DRN activity as it has been shown in thalamocortical connections (Crandall et al., 2015). In this scenario, given that the inhibitory interneurons in the DRN can track faster frequencies than 5-HT neurons (Jin et al., 2015) and that 5-HT neurons undergo 5-HT_{1a} autoreceptor mediated inhibition upon dendritic NMDA receptor activation (De Kock et al., 2006), a prolonged activation of mPFC afferents to the DRN may, in turn, produce inhibition of 5-HT neurons. Nevertheless, this would not exclude that other, less explored, patterns of cortical activity in different frequency ranges may tune 5-HT neurons subpopulations in different ways under more naturalistic patterns of activation and may be the focus of future research.

In addition, while it's well described that the PL/IL cortices produce a dense innervation over the DRN, the adjacent PR and IL

cortices exert opposite effects on fear conditioning (Giustino & Maren, 2015) as well as on avoidance behaviors and behavioral inhibition (Capuzzo et al., 2020). This striking contraposition in their functional role leaves open the possibility of different circuit motifs on their DRN innervation that could explain a putative excitatory or inhibitory effect and that should also be the focus of future research.

The cortical subpopulation of DRN-projecting neurons manipulated in adult Rbp4-Cre mice in this study presented collateral projections that were particularly dense onto the dorsomedial striatum (Fig. S5), a pathway that has been shown relevant for foraging decisions (Bari et al., 2019). While it has been shown that the cortico-striatal pathway is fully developed after P14 using Rbp4-Cre mice (Peixoto et al., 2016) and therefore unlikely to underlie the developmental differences observed in this study, we cannot rule out an impact of the absence of corticostriatal collaterals in the behavioral persistence decrease observed in Caspase treated mice.

The presence of parallel sub-systems in the DRN, with complementary projections either to the prefrontal cortex or to the amygdala and responsible for different behavioral responses has recently been reported (Ren et al., 2018). In our hands, mPFC-DRN descending neurons had very sparse collateralization to the amygdala (Fig. S5), while collaterals to the dorsal striatum or substantia nigra were abundant. This may suggest the presence of loops of preferential interconnectivity (mPFC→DRN/DRN→mPFC and mPFC→Amygdala/Amygdala→mPFC) as it has been shown for other cortical-subcortical loops (i.e. Young et al., 2021; Li et al., 2020), with different DRN subpopulations exerting specific neuromodulatory effects in either region (Ren et al., 2018).

To summarize, our results describe a process of late postnatal development of top-down mPFC afferents onto DRN causally linked to the emergence of behavioral persistence in the transition between adolescence and adulthood. This critical period of corticofugal axonal development may also represent a period of vulnerability for maladaptive development involved in the etiopathogenesis of psychiatric disorders (e.g. Rutter, 2007; Chen et al., 2018; Soiza-Reilly et al., 2019; Guirado et al., 2020).

Materials and Methods

Animal subjects: All experimental procedures were approved and performed in accordance with the Champalimaud Centre for the Unknown Ethics Committee guidelines and by the Portuguese Veterinary General Board (Direcção-Geral de Veterinária, approval 0421/000/000/2016). The mouse lines used in this study were obtained from the Mutant Mouse Resource and Research Center (MMRRC), Rbp4-Cre (stock number 031125-UCD), and from Jax Mice, Ai32(RCL-ChR2(H134R)/EYFP) (Stock number 012569) and Ai9(RCL-tdTomato) (7905). All of them were crossbred in-house for at least 10 generations prior to their use in our experiments. Mice were kept under a standard 12 h light/dark cycle with food and water ad libitum. Behavioral testing occurred during the light period.

Electrophysiological Recordings: For whole-cell recordings, 300 μm thick coronal slices containing the dorsal raphe were cut using a vibratome (Leica VT1200) in “ice cold” solution containing (in mM): 2.5 KCl, 1.25 NaH_2PO_4 , 26 NaHCO_3 , 10 D-glucose, 230 Sucrose, 0.5 CaCl_2 , 10 MgSO_4 , and bubbled with 5% CO_2 and 95% O_2 . Slices were recovered in ACSF containing (in mM): 127 NaCl, 2.5 KCl, 25 NaHCO_3 , 1.25 NaH_2PO_4 , 25 Glucose, 2 CaCl_2 , 1 MgCl_2 at 34 °C for 30 minutes and then kept in the same solution at room temperature until transferred to the recording chamber. In addition, 300 μM L-Tryptophan (Sigma) was added to the ACSF to maintain serotonergic tone in the ex vivo preparation as described elsewhere (Liu et al., 2005).

Patch recording pipettes (resistance 3-5 $\text{M}\Omega$) were filled with internal solution containing (in mM): 135 K-Gluconate, 10 HEPES, 10 Na-Phosphocreatine, 3 Na-L-Ascorbate, 4 MgCl_2 , 4 $\text{Na}_2\text{-ATP}$, and 0.4

Na-GTP. Data were acquired using a Multiclamp 700B amplifier and digitized at 10 kHz with a Digidata 1440a digitizer (both from Molecular Devices). Data was then either analyzed using Clampfit 10.7 Software (Molecular Devices, LLC) or imported into Matlab and analyzed with custom-written software.

Every voltage-clamp recording contained a 100ms test pulse of -10mV for offline calculation of access and series resistance to ensure the same recording quality across experiments.

The access resistance (R_a) was determined by measuring the amplitude of the current response to the command voltage step and the membrane resistance (R_m) as the difference between the baseline and the holding current in the steady state after the capacitive decay, by applying Ohm's law. The membrane time constant (τ) was determined by a single exponential fit of the decay phase in response to the square pulse. An approximation of the capacitance was obtained using the following formula:

$$\text{Tau} = \text{Access Resistance} * \text{Membrane Capacitance}$$

Only neurons with access resistance 1/20th lower than the membrane resistance were used for analysis. Only neurons with an access resistance lower than 25M Ω were considered for analysis. The access resistance was comparable between neurons recorded at different developmental stages (ANOVA, $F(2,101)=0.49$, $p=0.61$).

Neurons were recorded at a holding membrane voltage of -70mV, near the reversal potential of chloride (-68mV) and thus, optogenetically evoked responses correspond to AMPA-mediated currents.

To activate ChR2-expressing fibers, light from a 473-nm fiber-coupled laser (PSU-H-FDA, CNI Laser) was delivered at approximately 2mm from the sample to produce wide-field illumination of the recorded cell (Fig. 2A). TTL triggered pulses of light (10-ms duration; 10 mW measured at the fiber tip, which was located approximately 2 mm away from the sample) were delivered at the recording site at 0.1Hz of frequency. Neurons were recorded in the dorsomedial and lateral wings portions of the DRN, in two consecutive coronal slices per mouse between bregma levels ~4.3 and ~4.8 (Fig. S1).

To assess ChR2 evoked photocurrent in layer V somas, mPFC slices were obtained using the same slicing procedure. 1 μ M TTX was added to the bath to prevent escaped spikes in the voltage clamp recordings upon light activation.

Histology: Mice were deeply anesthetized with pentobarbital (Eutasil) and perfused transcardially with 4% paraformaldehyde (P6148, Sigma-Aldrich). The brain was removed from the skull, stored in 4% paraformaldehyde for two hours before being transferred to cryoprotectant solution (30% sucrose in PBS) until they sank. Sagittal sections (50 μ m) were cut with a freezing sliding microtome (SM2000, Leica).

For axonal quantification in Rbp4-ChR2, we performed anti-GFP immunostaining to enhance the intrinsic signal of the ChR2-fused EYFP reporter. We incubated overnight with the anti-GFP primary antibody at 4 degrees (1:1000, A-6455 Invitrogen, 0.1M PBS 0.3% tx100, 3% NGS). After abundant PBS washes, we incubated the secondary biotinylated anti-Rabbit antibody (711-065-152, Jackson IRL) for 2-4 hours in the same incubation solution at room temperature and finally, after PBS washes, slices were incubated in Alexa488 Streptavidin for 2-4 hours in

the same incubation solution at room temperature (S32354, Invitrogen). After final PBS washes, slices were mounted and covered with FluoroGel mounting medium (17985-10, Electron Microscopy Sciences) for posterior imaging.

Image acquisition and analysis: Histological sections were imaged with a Zeiss LSM 710 confocal laser scanning microscope using 10x and 25x magnification objectives.

To quantify the axonal density, images containing the DRN or the nucleus accumbens were background subtracted and binarized using constant thresholds in Fiji. After thresholding, binary images were imported into Matlab and analyzed with custom written software. Images were sampled every 100um across the Y axis. The intersections between binary axons and these sampling lines across the Y axis were counted and averaged in bins of 100um to estimate axonal density.

Quantification of layer V soma densities was obtained in histological slices of Rbp4-tdTomato mice. Confocal images were imported into matlab, and fluorescent somas were detected using the image analysis toolbox of Matlab inside a defined region of interest containing the mPFC (PL/IL). The number of somas was then divided by the area of the ROI to obtain the density of neurons.

Stereotaxic surgeries and virus injection: Animals were anesthetized with isoflurane (2% induction and 0.5 - 1% for maintenance) and placed in a motorized computer-controlled Stoelting stereotaxic instrument with mouse brain atlas integration and real-time surgery probe visualization in the atlas space (Neurostar, Sindelfingen, Germany; <https://www.neurostar.de>). Antibiotic (Enrofloxacin, 2.5-5 mg/Kg, S.C.), pain killer (Buprenorphine, 0.1 mg/Kg, S.C.), and local

anesthesia over the scalp (0.2 ml, 2%Lidocaine, S.C.) were administered before incising the scalp. Virus injection (experiment group: AAV2retro-flex-EF1A-taCasp3-TEVp; control group: AAV2retro-flex-hSyn-tdTomato) was targeted to DRN at the following coordinates: -4.7 mm AP, 0.0 mm ML, and 3.1 mm DV. The vertical stereotaxic arm was tilted 32 degrees caudally to reach the target avoiding Superior sagittal sinus and Transverse sinuses. Target coordinates were adjusted as follows: -6.64 mm AP, 0.0 mm ML, and -4.02 mm DV. To infect a larger volume of the DRN with the virus, we performed six injections of 0.2 uL using two entry points along the AP axis (-6.54 and -6.74) and 3 depths along the DV axis(-4.02, -3.92, and -3.82). The incision was then closed using tissue adhesive (VETBONDTM/MC, 3M, No. 1469SB). Mice were monitored until recovery from the surgery and returned to their homecages, where they were housed individually. Behavioral testing started at least 1 week after surgery to allow for recovery.

Behavioral testing:

-SETUP: The behavioral box consisted of 1 back-wall (16x219 cm), 2 side-walls (16.7x219 cm), and 2 front-walls (10x219 cm,140-degrees angle between them), made of white acrylic (0.5 cm thick) and a transparent acrylic lead. A camera (ELP camera, ELP-USBFHD01M-L180) was mounted on top of the ceiling for monitoring purposes. Each front wall had a nose-poke port equipped with an infrared emitter/sensor pairs to report port entry and exit times (model 007120.0002, Island motion corporation) and a water valve for water delivery (LHDA1233115H, The Lee Company, Westbrook, CT). An internal white acrylic wall (8cm) separates the two nose-poke ports forcing the animals to walk around it to travel between ports. All signals

from sensors were processed by Arduino Mega 2560 microcontroller board (Arduino, Somerville, US), and outputs from the Arduino Mega 2560 microcontroller board were implemented to control water delivery in drops of 4 μ l. Arduino Mega 2560 microcontroller was connected to the sensors and controllers through an Arduino Mega 2560 adaptor board developed by the Champalimaud Foundation Scientific Hardware Platform.

-TASK: Subjects have to probe two foraging sites (nose-poke ports, for mice, or virtual magic wands, for humans) to obtain rewards (4 μ l water drops, for mice, or virtual points for humans). At any given time, only one of the sites is active and, when probed, delivers a reward with a fixed 90% probability (P_{REW}). Each attempt also triggers a fixed 90% probability of transition (P_{TRS}) to inactivate the current foraging site and activate the other. These transitions are not cued; thus, subjects are required to alternate probing the current site and traveling to the other to track the hidden active state and obtain rewards. In this work, we focus on assessing differences in the baseline patience/impulsivity, measured as the ability to withhold adverse outcomes. Therefore naive subjects were only tested once.

-BEHAVIORAL PROTOCOL FOR MICE: Five days before testing, water dispensers were removed from the animals' home cages, and their weights were recorded. In the following days, progressively less water (1000 μ l, 800 μ l, 600 μ l) was given in a metal dish inside the home cage. Weight loss was monitored every day before water delivery, and no animal lost more than 20% of their body weight. On the fifth day of water deprivation, animals were weighed and introduced to the behavioral box. A small quantity of water was present at the start of the session to stimulate the mice to probe the nose-ports. Sessions lasted a

minimum of one hour. By that time, if animals did not perform at least 30 trials, the session was extended for thirty more minutes.

Animals were handled during water deprivation to reduce stress levels, but they were completely naive about the task environment and functioning on the testing day. One juvenile female mouse was excluded from the experiment batch before the task assessment because of congenital blindness. One caspase adult mouse was excluded after the task assessment because of abnormal behavior. Rather than nose poking to seek water this animal spent most of the task time biting the nose port, to anomalous levels.

Data and statistical analysis: Behavioral data analysis was performed using custom-written scripts in Julia-1.4.1. Behavioral results were represented as median $\pm$ 95% confidence intervals, and statistical significance was accepted for p -values < 0.05 . The statistical analysis was done in Julia-1.4.1 with the HypothesisTests and MixedModels existing packages. Unpaired group differences (proportion of poking time per trial and proportion of premature leaving) were tested by performing the Mann-Whitney U test, followed by the effect size calculation if $p < 0.05$. The effect of a specific factor on the probability of leaving was tested by applying logistic regression on a generalized linear mixed-effects model (GLMM), using a Bernoulli distribution for the dependent variable and a Logit link function. First, to assess the significance of the estimated coefficients, we calculated their 95% CI by performing a parametric bootstrap of 1000 samples. Only factors whose CI did not include 0 were considered to be significantly affecting the probability of leaving. Next, to validate the relevance of the experimental manipulation (age or virus), we compared nested models (a general model and a special case model, excluding or including the experimental factor, respectively) using a likelihood ratio test: chi-squared test on the

difference of the deviance of the two nested models, with degrees of freedom equal to the difference in degrees of freedom between the general model (lacking the predictor) and its special case (with the predictor of interest). For each analysis, we report the median and 95% CI of the median for the groups of interest, followed by the test statistics. We use Wilkinson annotation to describe the models with denoting random effects.

Electrophysiological and histological results were analyzed with Matlab and Graphpad Software. Normality of the residuals was tested with the D'Agostino-Pearson omnibus K2 test. When normally distributed, either a t-test, one-way ANOVA or repeated measures ANOVA were performed to compare groups at different developmental phases. In the cases where residuals were not normally distributed, we performed a Mann-Whitney or Kruskal Wallis test to assess significance. For testing differences in connection probability, a Chi-square test was performed. Finally, a Kolmogorov Smirnov test was performed to compare the neuronal density distribution between Caspase treated animals and tdTomato expressing controls. Error bar plots represent mean $\pm$ SEM. Significance was noted as * $p < 0.05$.

References

- Aguillon-Rodriguez, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., Cazettes, F., Chapuis, G., Churchland, A. K., Dan, Y., Dewitt, E., Faulkner, M., Forrest, H., Haetzel, L., Häusser, M., Hofer, S. B., Hu, F., Khanal, A., Krasniak, C., Laranjeira, I., ... Zador, A. M. (2021). Standardized and reproducible measurement of decision-making in mice. *ELife*, 10. <https://doi.org/10.7554/eLife.63711>
- Alexander-Bloch, A., Giedd, J. N., & Bullmore, E. (2013). Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience*, 14(5), 322–336. <https://doi.org/10.1038/nrn3465>
- Arruda-Carvalho, M., Wu, W. C., Cummings, K. A., & Clem, R. L. (2017). Optogenetic examination of prefrontal-amygdala synaptic development. *Journal of Neuroscience*, 37(11), 2976–2985. <https://doi.org/10.1523/JNEUROSCI.3097-16.2017>
- Ashwell, R., & Ito, R. (2014). Excitotoxic lesions of the infralimbic, but not prelimbic cortex facilitate reversal of appetitive discriminative context conditioning: the role of the infralimbic cortex in context generalization. *Frontiers in Behavioral Neuroscience*, 8(FEB), 63. <https://doi.org/10.3389/fnbeh.2014.00063>
- Bari, B. A., Grossman, C. D., Lubin, E. E., Rajagopalan, A. E., Cressy, J. I., & Cohen, J. Y. (2019). Stable Representations of Decision Variables for Flexible Behavior. *Neuron*, 103(5), 922-933.e7. <https://doi.org/10.1016/j.neuron.2019.06.001>
- Bell, M. R. (2018). Comparing Postnatal Development of Gonadal Hormones and Associated Social Behaviors in Rats, Mice, and Humans. *Endocrinology*, 159(7), 2596–2613. <https://doi.org/10.1210/EN.2018-00220>
- Berlin, H. A. (2004). Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. *Brain*, 127(5), 1108–1126. <https://doi.org/10.1093/brain/awh135>
- Bitzenhofer, S. H., Pöplau, J. A., Chini, M., Marquardt, A., & Hanganu-Opatz, I. L. (2021). A transient developmental increase in prefrontal activity alters network maturation and causes cognitive dysfunction in adult mice. *Neuron*, 109(8), 1350-1364.e6. <https://doi.org/10.1016/j.neuron.2021.02.011>
- Capuzzo, G., & Floresco, S. B. (2020). Prelimbic and Infralimbic Prefrontal Regulation of Active and Inhibitory Avoidance and Reward-Seeking. *The Journal of Neuroscience*, 40(24), 4773–4787. <https://doi.org/10.1523/JNEUROSCI.0414-20.2020>
- Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., Jonides, J., Berman, M. G., Wilson, N. L., Teslovich, T., Glover, G.,

Zayas, V., Mischel, W., & Shoda, Y. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences*, 108(36), 14998–15003. <https://doi.org/10.1073/pnas.1108561108>

Celada, P., Puig, M. V., Casanovas, J. M., Guillazo, G., & Artigas, F. (2001). Control of Dorsal Raphe Serotonergic Neurons by the Medial Prefrontal Cortex: Involvement of Serotonin-1A, GABAA, and Glutamate Receptors. *The Journal of Neuroscience*, 21(24), 1–13. <https://doi.org/10.1523/JNEUROSCI.21-24-09917.2001>

Challis, C., Beck, S. G., & Berton, O. (2014). Optogenetic modulation of descending prefrontocortical inputs to the dorsal raphe bidirectionally bias socioaffective choices after social defeat. *Frontiers in Behavioral Neuroscience*, 8(FEB), 1–14. <https://doi.org/10.3389/fnbeh.2014.00043>

Chen, F., Ke, J., Qi, R., Xu, Q., Zhong, Y., Liu, T., Li, J., Zhang, L., & Lu, G. (2018). Increased Inhibition of the Amygdala by the mPFC may Reflect a Resilience Factor in Post-traumatic Stress Disorder: A Resting-State fMRI Granger Causality Analysis. *Frontiers in Psychiatry*, 9, 516. <https://doi.org/10.3389/fpsy.2018.00516>

Chini, M., & Hanganu-Opatz, I. L. (2021). Prefrontal Cortex Development in Health and Disease: Lessons from Rodents and Humans. *Trends in Neurosciences*, 44(3), 227–240. <https://doi.org/10.1016/j.tins.2020.10.017>

Ciamarelli, E., De Luca, F., Kwan, D., Mok, J., Bianconi, F., Knyagnytska, V., Craver, C., Green, L., Myerson, J., & Rosenbaum, R. S. (2021). The role of ventromedial prefrontal cortex in reward valuation and future thinking during intertemporal choice. *ELife*, 10, 1–17. <https://doi.org/10.7554/eLife.67387>

Dalley, J. W., & Robbins, T. W. (2017). Fractionating impulsivity : neuropsychiatric implications. *Nature Publishing Group*, 18(3), 158–171. <https://doi.org/10.1038/nrn.2017.8>

De Kock, C. P. J., Cornelisse, L. N., Burnashev, N., Lodder, J. C., Timmerman, A. J., Couey, J. J., Mansvelder, H. D., & Brussaard, A. B. (2006). NMDA receptors trigger neurosecretion of 5-HT within dorsal raphe nucleus of the rat in the absence of action potential firing. *The Journal of Physiology*, 577(3), 891–905. <https://doi.org/10.1113/jphysiol.2006.115311>

Dincheva, I., Drysdale, A. T., Hartley, C. A., Johnson, D. C., Jing, D., King, E. C., Ra, S., Gray, J. M., Yang, R., DeGruccio, A. M., Huang, C., Cravatt, B. F., Glatt, C. E., Hill, M. N., Casey, B. J., & Lee, F. S. (2015). FAAH genetic variation enhances fronto-amygdala function in mouse and human. *Nature Communications*, 6(1), 6395. <https://doi.org/10.1038/ncomms7395>

Durston, S., & Casey, B. J. (2006). What have we learned about cognitive development from neuroimaging? *Neuropsychologia*, 44(11), 2149–2157. <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2005.10.010>

- Fair, D. A., Cohen, A. L., Power, J. D., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2009). Functional Brain Networks Develop from a “Local to Distributed” Organization. *PLoS Computational Biology*, 5(5), e1000381. <https://doi.org/10.1371/journal.pcbi.1000381>
- Fellows, L. K. (2006). Deciding how to decide: Ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. *Brain*, 129(4), 944–952. <https://doi.org/10.1093/brain/awl017>
- Ferguson, B. R., & Gao, W.-J. (2015). Development of thalamocortical connections between the mediodorsal thalamus and the prefrontal cortex and its implication in cognition. *Frontiers in Human Neuroscience*, 8(JAN). <https://doi.org/10.3389/fnhum.2014.01027>
- Fonseca, M. S., Murakami, M., & Mainen, Z. F. (2015). Activation of dorsal raphe serotonergic neurons promotes waiting but is not reinforcing. *Current Biology*, 25(3), 306–315. <https://doi.org/10.1016/j.cub.2014.12.002>
- Geddes, S. D., Assadzada, S., Lemelin, D., Sokolovski, A., Bergeron, R., Haj-Dahmane, S., & Béique, J.-C. (2016). Target-specific modulation of the descending prefrontal cortex inputs to the dorsal raphe nucleus by cannabinoids. *Proceedings of the National Academy of Sciences*, 113(19), 5429–5434. <https://doi.org/10.1073/pnas.1522754113>
- Gee, D. G., Fetcho, R. N., Jing, D., Li, A., Glatt, C. E., Drysdale, A. T., Cohen, A. O., Dellarco, D. V., Yang, R. R., Dale, A. M., Jernigan, T. L., Lee, F. S., & Casey, B. J. (2016). Individual differences in frontolimbic circuitry and anxiety emerge with adolescent changes in endocannabinoid signaling across species. *Proceedings of the National Academy of Sciences*, 113(16), 4500–4505. <https://doi.org/10.1073/pnas.1600013113>
- Giustino, T. F., & Maren, S. (2015). The Role of the Medial Prefrontal Cortex in the Conditioning and Extinction of Fear. *Frontiers in Behavioral Neuroscience*, 9(NOVEMBER). <https://doi.org/10.3389/fnbeh.2015.00298>
- Gonçalves, L., Nogueira, M. I., Shammah-Lagnado, S. J., & Metzger, M. (2009). Prefrontal afferents to the dorsal raphe nucleus in the rat. *Brain Research Bulletin*, 78(4–5), 240–247. <https://doi.org/10.1016/j.brainresbull.2008.11.012>
- Gorelova, N., & Yang, C. . (1996). The course of neural projection from the prefrontal cortex to the nucleus accumbens in the rat. *Neuroscience*, 76(3), 689–706. [https://doi.org/10.1016/S0306-4522\(96\)00380-6](https://doi.org/10.1016/S0306-4522(96)00380-6)
- Guirado, R., Perez-Rando, M., Ferragud, A., Gutierrez-Castellanos, N., Umemori, J., Carceller, H., Nacher, J., & Castillo-Gómez, E. (2020). A Critical Period for Prefrontal Network Configurations Underlying Psychiatric Disorders and Addiction. *Frontiers in Behavioral Neuroscience*, 14, 51. <https://doi.org/10.3389/fnbeh.2020.00051>

- Hamani, C., Diwan, M., Macedo, C. E., Brandão, M. L., Shumake, J., Gonzalez-Lima, F., Raymond, R., Lozano, A. M., Fletcher, P. J., & Nobrega, J. N. (2010). Antidepressant-Like Effects of Medial Prefrontal Cortex Deep Brain Stimulation in Rats. *Biological Psychiatry*, 67(2), 117–124. <https://doi.org/10.1016/J.BIOPSYCH.2009.08.025>
- Hvoslef-Eide, M., Nilsson, S. R. O., Hailwood, J. M., Robbins, T. W., Saksida, L. M., Mar, A. C., & Bussey, T. J. (2018). Effects of anterior cingulate cortex lesions on a continuous performance task for mice. *Brain and Neuroscience Advances*, 2, 239821281877296. <https://doi.org/10.1177/2398212818772962>
- Hwang, K., Velanova, K., & Luna, B. (2010). Strengthening of Top-Down Frontal Cognitive Control Networks Underlying the Development of Inhibitory Control: A Functional Magnetic Resonance Imaging Effective Connectivity Study. *Journal of Neuroscience*, 30(46), 15535–15545. <https://doi.org/10.1523/JNEUROSCI.2825-10.2010>
- Itami, S., & Uno, H. (2002). Orbitofrontal cortex dysfunction in attention-deficit hyperactivity disorder revealed by reversal and extinction tasks. *NeuroReport*, 13(18), 2453–2457. <https://doi.org/10.1097/00001756-200212200-00016>
- Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, 345(March), 12–26. <https://doi.org/10.1016/j.neuroscience.2016.03.021>
- Jin, Y., Luo, B., Su, Y.-Y., Wang, X.-X., Chen, L., Wang, M., Wang, W.-W., & Chen, L. (2015). Sodium Salicylate Suppresses GABAergic Inhibitory Activity in Neurons of Rodent Dorsal Raphe Nucleus. *PLOS ONE*, 10(5), e0126956. <https://doi.org/10.1371/journal.pone.0126956>
- Johnson, C., & Wilbrecht, L. (2011). Juvenile mice show greater flexibility in multiple choice reversal learning than adults. *Developmental Cognitive Neuroscience*, 1(4), 540–551. <https://doi.org/10.1016/j.dcn.2011.05.008>
- Klune, C. B., Jin, B., & DeNardo, L. A. (2021). Linking mPFC circuit maturation to the developmental regulation of emotional memory and cognitive flexibility. *ELife*, 10, 1–33. <https://doi.org/10.7554/eLife.64567>
- Kolk, S. M., & Rakic, P. (2021). Development of prefrontal cortex. *Neuropsychopharmacology*, 1–17. <https://doi.org/10.1038/s41386-021-01137-9>
- KW, M., K, M., & K, D. (2012). Activation of dorsal raphe serotonin neurons is necessary for waiting for delayed rewards. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 32(31), 10451–10457. <https://doi.org/10.1523/JNEUROSCI.0915-12.2012>
- Larsen, B., & Luna, B. (2018). Adolescence as a neurobiological critical period for the development of higher-order cognition. *Neuroscience & Biobehavioral Reviews*, 94, 179–195. <https://doi.org/10.1016/j.neubiorev.2018.09.005>

- Laviola, G., Macrì, S., Morley-Fletcher, S., & Adriani, W. (2003). Risk-taking behavior in adolescent mice: psychobiological determinants and early epigenetic influence. *Neuroscience & Biobehavioral Reviews*, 27(1–2), 19–31. [https://doi.org/10.1016/S0149-7634\(03\)00006-X](https://doi.org/10.1016/S0149-7634(03)00006-X)
- Leone, D. P., Heavner, W. E., Ferenczi, E. A., Dobрева, G., Huguenard, J. R., Grosschedl, R., & McConnell, S. K. (2015). *Satb2* Regulates the Differentiation of Both Callosal and Subcerebral Projection Neurons in the Developing Cerebral Cortex. *Cerebral Cortex*, 25(10), 3406–3419. <https://doi.org/10.1093/cercor/bhu156>
- Li, B., Nguyen, T. P., Ma, C., & Dan, Y. (2020). Inhibition of impulsive action by projection-defined prefrontal pyramidal neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 117(29), 17278–17287. <https://doi.org/10.1073/PNAS.2000523117/-DCSUPPLEMENTAL>
- Li, Z., Chen, Z., Fan, G., Li, A., Yuan, J., & Xu, T. (2018). Cell-Type-Specific Afferent Innervation of the Nucleus Accumbens Core and Shell. *Frontiers in Neuroanatomy*, 12. <https://doi.org/10.3389/fnana.2018.00084>
- Lottem, E., Banerjee, D., Vertechi, P., Sarra, D., Lohuis, M. oude, & Mainen, Z. F. (2018). Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature Communications*, 9(1), 1000. <https://doi.org/10.1038/s41467-018-03438-y>
- Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., & Chahal, R. (2015). An Integrative Model of the Maturation of Cognitive Control. *Annual Review of Neuroscience*, 38(1), 151–170. <https://doi.org/10.1146/annurev-neuro-071714-034054>
- Luna, B., Thulborn, K. R., Munoz, D. P., Merriam, E. P., Garver, K. E., Minshew, N. J., Keshavan, M. S., Genovese, C. R., Eddy, W. F., & Sweeney, J. A. (2001). Maturation of widely distributed brain function subserves cognitive development. *NeuroImage*, 13(5), 786–793. <https://doi.org/10.1006/nimg.2000.0743>
- Maier, S. F. (2015). Behavioral control blunts reactions to contemporaneous and future adverse events: Medial prefrontal cortex plasticity and a corticostriatal network. *Neurobiology of Stress*, 1(1), 12–22. <https://doi.org/10.1016/j.ynstr.2014.09.003>
- Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of Gratification in Children. ii, 21–26.
- Miyazaki, K. K. W., Miyazaki, K. K. W., Tanaka, K. F., Yamanaka, A., Takahashi, A., Tabuchi, S., & Doya, K. (2014). Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. *Current Biology*, 24(17), 2033–2040. <https://doi.org/10.1016/j.cub.2014.07.041>
- Miyazaki, K., Miyazaki, K. W., Yamanaka, A., Tokuda, T., Tanaka, K. F., & Doya, K. (2018). Reward probability and timing uncertainty alter the effect of dorsal

raphe serotonin neurons on patience. *Nature Communications*, 9(1).
<https://doi.org/10.1038/s41467-018-04496-y>

Miyazaki, K. W., Miyazaki, K., & Doya, K. (2011). Activation of the central serotonergic system in response to delayed but not omitted rewards. *European Journal of Neuroscience*, 33(1), 153–160.
<https://doi.org/10.1111/j.1460-9568.2010.07480.x>

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693–2698. <https://doi.org/10.1073/pnas.1010076108>

Morris, D. W., & Davidson, D. L. (2000). Optimally Foraging Mice Match Patch Use with Habitat Differences in Fitness. *Ecology*, 81(8), 2061.
<https://doi.org/10.2307/177095>

Muir, J. L., Everitt, B. J., & Robbins, T. W. (1996). The Cerebral Cortex of the Rat and Visual Attentional Function: Dissociable Effects of Medial Frontal, Cingulate, Anterior Dorsolateral, and Parietal Cortex Lesions on a Five-Choice Serial Reaction Time Task. *Cerebral Cortex*, 6(3), 470–481.
<https://doi.org/10.1093/cercor/6.3.470>

Muir, J., Lopez, J., & Bagot, R. C. (2019). Wiring the depressed brain: optogenetic and chemogenetic circuit interrogation in animal models of depression. *Neuropsychopharmacology*, 44(6), 1013–1026.
<https://doi.org/10.1038/s41386-018-0291-6>

Murakami, M., Shteingart, H., Loewenstein, Y., & Mainen, Z. F. (2017). Distinct Sources of Deterministic and Stochastic Components of Action Timing Decisions in Rodent Frontal Cortex. *Neuron*, 94(4), 908–919.e7.
<https://doi.org/10.1016/j.neuron.2017.04.040>

Nagy, Z., Westerberg, H., & Klingberg, T. (2004). Maturation of White Matter is Associated with the Development of Cognitive Functions during Childhood. *Journal of Cognitive Neuroscience*, 16(7), 1227–1233.
<https://doi.org/10.1162/0898929041920441>

Nakayama, H., Ibañez-Tallon, I., & Heintz, N. (2018). Cell-Type-Specific Contributions of Medial Prefrontal Neurons to Flexible Behaviors. *The Journal of Neuroscience*, 38(19), 4490–4504.
<https://doi.org/10.1523/JNEUROSCI.3537-17.2018>

Narayanan, N. S., Horst, N. K., & Laubach, M. (2006). Reversible inactivations of rat medial prefrontal cortex impair the ability to wait for a stimulus. *Neuroscience*, 139(3), 865–876.
<https://doi.org/10.1016/j.neuroscience.2005.11.072>

Narayanan, N. S., Cavanagh, J. F., Frank, M. J., & Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. *Nature Neuroscience*, 16(12), 1888–1895. <https://doi.org/10.1038/nn.3549>

Nishitani, N., Nagayasu, K., Asaoka, N., Yamashiro, M., Andoh, C., Nagai, Y., Kinoshita, H., Kawai, H., Shibui, N., Liu, B., Hewinson, J., Shirakawa, H., Nakagawa, T., Hashimoto, H., Kasparov, S., & Kaneko, S. (2019). Manipulation of dorsal raphe serotonergic neurons modulates active coping to inescapable stress and anxiety-related behaviors in mice and rats. *Neuropsychopharmacology*, 44(4), 721–732. <https://doi.org/10.1038/s41386-018-0254-y>

Ohmura, Y., Tsutsui-Kimura, I., Sasamori, H., Nebuka, M., Nishitani, N., Tanaka, K. F., Yamanaka, A., & Yoshioka, M. (2020). Different roles of distinct serotonergic pathways in anxiety-like behavior, antidepressant-like, and anti-impulsive effects. *Neuropharmacology*, 167(July), 107703. <https://doi.org/10.1016/j.neuropharm.2019.107703>

Peixoto, R. T., Wang, W., Croney, D. M., Kozorovitskiy, Y., & Sabatini, B. L. (2016). Early hyperactivity and precocious maturation of corticostriatal circuits in Shank3B^{-/-} mice. *Nature Neuroscience*, 19(5), 716–724. <https://doi.org/10.1038/nn.4260>

Petreaanu, L., Mao, T., Sternson, S. M., & Svoboda, K. (2009). The subcellular organization of neocortical excitatory connections. *Nature*, 457(7233), 1142–1145. <https://doi.org/10.1038/nature07709>

Phillipson, O. T., & Griffiths, A. C. (1985). The topographic order of inputs to nucleus accumbens in the rat. *Neuroscience*, 16(2), 275–296. [https://doi.org/10.1016/0306-4522\(85\)90002-8](https://doi.org/10.1016/0306-4522(85)90002-8)

Pittaras, E., Rabat, A., & Granon, S. (2020). The Mouse Gambling Task: Assessing Individual Decision-making Strategies in Mice. *Bio-Protocol*, 10(1). <https://doi.org/10.21769/BIOPROTOCOL.3479>

Pollak Dorocic, I., Fürth, D., Xuan, Y., Johansson, Y., Pozzi, L., Silberberg, G., Carlén, M., & Meletis, K. (2014). A Whole-Brain Atlas of Inputs to Serotonergic Neurons of the Dorsal and Median Raphe Nuclei. *Neuron*, 83(3), 663–678. <https://doi.org/10.1016/j.neuron.2014.07.002>

Puig, M. V., & Gullledge, A. T. (2011). Serotonin and prefrontal cortex function: neurons, networks, and circuits. *Molecular Neurobiology*, 44(3), 449–464. <https://doi.org/10.1007/S12035-011-8214-0>

Ren, J., Friedmann, D., Xiong, J., Liu, C. D., Ferguson, B. R., Weerakkody, T., DeLoach, K. E., Ran, C., Pun, A., Sun, Y., Weissbourd, B., Neve, R. L., Huguenard, J., Horowitz, M. A., & Luo, L. (2018). Anatomically Defined and Functionally Distinct Dorsal Raphe Serotonin Sub-systems. *Cell*, 175(2), 472–487.e20. <https://doi.org/10.1016/j.cell.2018.07.043>

Roberts, A. C., Robbins, T. W., & Weiskrantz, L. (1998). The Prefrontal Cortex: Executive and Cognitive Functions. In A. C. Roberts, T. W. Robbins, & L. Weiskrantz (Eds.), *The Prefrontal Cortex Executive and Cognitive Functions*. Oxford University Press.

<https://doi.org/10.1093/ACPROF:OSO/9780198524410.001.0001>

Rudebeck, P. H., Saunders, R. C., Prescott, A. T., Chau, L. S., & Murray, E. A. (2013). Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nature Neuroscience*, 16(8), 1140–1145.

<https://doi.org/10.1038/nn.3440>

Rutter, M., Beckett, C., Castle, J., Colvert, E., Kreppner, J., Mehta, M., Stevens, S., & Sonuga-Barke, E. (2007). Effects of profound early institutional deprivation: An overview of findings from a UK longitudinal study of Romanian adoptees. *European Journal of Developmental Psychology*, 4(3), 332–350.

<https://doi.org/10.1080/17405620701401846>

Sackett, D. A., Moschak, T. M., & Carelli, R. M. (2019). Prelimbic Cortical Neurons Track Preferred Reward Value and Reflect Impulsive Choice during Delay Discounting Behavior. *The Journal of Neuroscience*, 39(16), 3108–3118.

<https://doi.org/10.1523/JNEUROSCI.2532-18.2019>

Sakurai, T., & Gamo, N. J. (2019). Cognitive functions associated with developing prefrontal cortex during adolescence and developmental neuropsychiatric disorders. *Neurobiology of Disease*, 131, 104322.

<https://doi.org/10.1016/j.nbd.2018.11.007>

Sanchez-Roige, S., Peña-Oliver, Y., & Stephens, D. N. (2012). Measuring impulsivity in mice: the five-choice serial reaction time task.

Psychopharmacology, 219(2), 253–270.

<https://doi.org/10.1007/s00213-011-2560-5>

Sasamori, H., Ohmura, Y., Kubo, T., Yoshida, T., & Yoshioka, M. (2018). Assessment of impulsivity in adolescent mice: A new training procedure for a 3-choice serial reaction time task. *Behavioural Brain Research*, 343, 61–70.

<https://doi.org/10.1016/j.bbr.2018.01.014>

Schweighofer, N., Bertin, M., Shishida, K., Okamoto, Y., Tanaka, S. C., Yamawaki, S., & Doya, K. (2008). Low-Serotonin Levels Increase Delayed Reward Discounting in Humans. *Journal of Neuroscience*, 28(17), 4528–4532.

<https://doi.org/10.1523/JNEUROSCI.4982-07.2008>

Sercombe, H. (2014). Risk, adaptation and the functional teenage brain. *Brain and Cognition*, 89, 61–69. <https://doi.org/10.1016/j.bandc.2014.01.001>

Soiza-Reilly, M., Meye, F. J., Olusakin, J., Telley, L., Petit, E., Chen, X., Mameli, M., Jabaudon, D., Sze, J. Y., & Gaspar, P. (2019). SSRIs target prefrontal to raphe circuits during development modulating synaptic connectivity and emotional behavior. *Molecular Psychiatry*, 24(5), 726–745.

<https://doi.org/10.1038/s41380-018-0260-9>

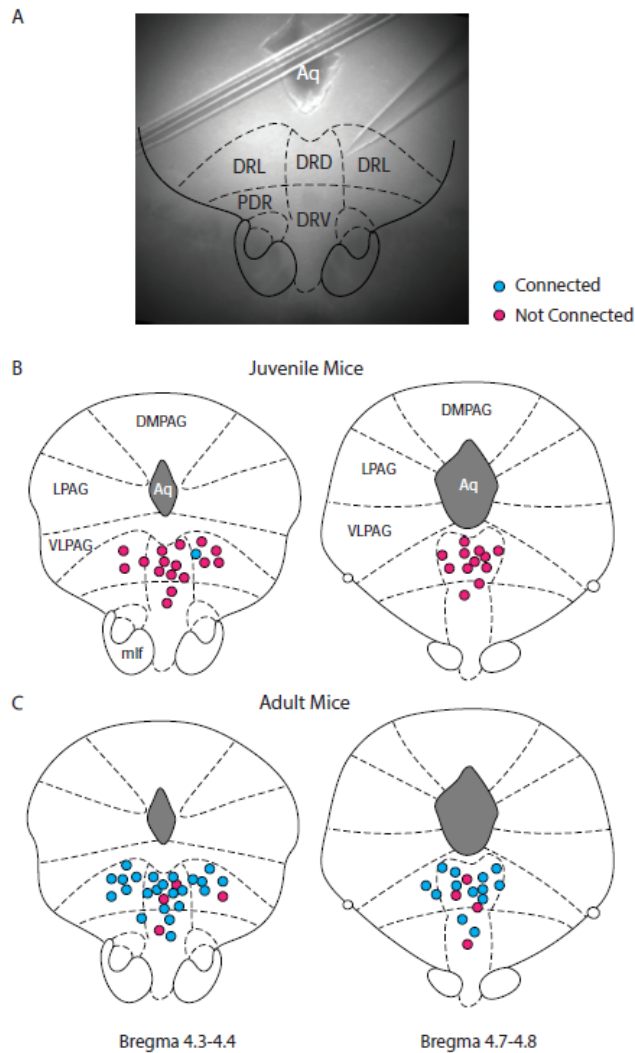
- Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience*, 2(10), 859–861.
<https://doi.org/10.1038/13154>
- Spear, L. P. (2000). Neurobehavioral Changes in Adolescence. *Current Directions in Psychological Science*, 9(4), 111–114.
<https://doi.org/10.1111/1467-8721.00072>
- Sreijic, L. R., Hamani, C., & Hutchison, W. D. (2015). High-frequency stimulation of the medial prefrontal cortex decreases cellular firing in the dorsal raphe. *European Journal of Neuroscience*, 41(9), 1219–1226.
<https://doi.org/10.1111/ejn.12856>
- Tooley, U. A., Bassett, D. S., & Mackey, A. P. (2021). Environmental influences on the pace of brain development. *Nature Reviews Neuroscience*, 22(6), 372–384. <https://doi.org/10.1038/s41583-021-00457-5>
- Ueda, S., Niwa, M., Hioki, H., Sohn, J., Kaneko, T., Sawa, A., & Sakurai, T. (2015). Sequence of Molecular Events during the Maturation of the Developing Mouse Prefrontal Cortex. *Molecular Neuropsychiatry*, 1(2), 94–104.
<https://doi.org/10.1159/000430095>
- van der Veen, B., Kapaniaiah, S. K. T., Kilonzo, K., Steele-Perkins, P., Jendryka, M. M., Schulz, S., Tasic, B., Yao, Z., Zeng, H., Akam, T., Nicholson, J. R., Liss, B., Nissen, W., Pekcec, A., & Kätzel, D. (2021). Control of impulsivity by Gi-protein signalling in layer-5 pyramidal neurons of the anterior cingulate cortex. *Communications Biology*, 4(1), 662.
<https://doi.org/10.1038/s42003-021-02188-w>
- Velanova, K., Wheeler, M. E., & Luna, B. (2008). Maturation Changes in Anterior Cingulate and Frontoparietal Recruitment Support the Development of Error Processing and Inhibitory Control. *Cerebral Cortex*, 18(11), 2505–2522.
<https://doi.org/10.1093/cercor/bhn012>
- Verharen, J. P. H., den Ouden, H. E. M., Adan, R. A. H., & Vanderschuren, L. J. M. J. (2020). Modulation of value-based decision making behavior by subregions of the rat prefrontal cortex. *Psychopharmacology*, 237(5), 1267–1280. <https://doi.org/10.1007/s00213-020-05454-7>
- Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., Oude Lohuis, M. N., & Mainen, Z. F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. *Neuron*, 106(1), 166–176.e6. <https://doi.org/10.1016/j.neuron.2020.01.017>
- Warden, M. R., Selimbeyoglu, A., Mirzabekov, J. J., Lo, M., Thompson, K. R., Kim, S. Y., Adhikari, A., Tye, K. M., Frank, L. M., & Deisseroth, K. (2012). A prefrontal cortex-brainstem neuronal projection that controls response to behavioural challenge. *Nature*, 492(7429), 428–432.
<https://doi.org/10.1038/nature11617>

- Warthen, D. M., Lambeth, P. S., Ottolini, M., Shi, Y., Barker, B. S., Gaykema, R. P., Newmyer, B. A., Joy-Gaba, J., Ohmura, Y., Perez-Reyes, E., Güler, A. D., Patel, M. K., & Scott, M. M. (2016). Activation of Pyramidal Neurons in Mouse Medial Prefrontal Cortex Enhances Food-Seeking Behavior While Reducing Impulsivity in the Absence of an Effect on Food Intake. *Frontiers in Behavioral Neuroscience*, 10(MAR). <https://doi.org/10.3389/fnbeh.2016.00063>
- Weissbourd, B., Ren, J., DeLoach, K. E., Guenther, C. J., Miyamichi, K., & Luo, L. (2014). Presynaptic Partners of Dorsal Raphe Serotonergic and GABAergic Neurons. *Neuron*, 83(3), 645–662. <https://doi.org/10.1016/j.neuron.2014.06.024>
- Winstanley, C. A., & Floresco, S. B. (2016). Deciphering Decision Making: Variation in Animal Models of Effort- and Uncertainty-Based Choice Reveals Distinct Neural Circuitries Underlying Core Cognitive Processes. *Journal of Neuroscience*, 36(48), 12069–12079. <https://doi.org/10.1523/JNEUROSCI.1713-16.2016>
- Wong, M. M., Nigg, J. T., Zucker, R. A., Puttler, L. I., Fitzgerald, H. E., Jester, J. M., Glass, J. M., & Adams, K. (2006). Behavioral Control and Resiliency in the Onset of Alcohol and Illicit Drug Use: A Prospective Study From Preschool to Adolescence. *Child Development*, 77(4), 1016–1033. <https://doi.org/10.1111/j.1467-8624.2006.00916.x>
- Yang, C. F., Chiang, M. C., Gray, D. C., Prabhakaran, M., Alvarado, M., Juntti, S. A., Unger, E. K., Wells, J. A., & Shah, N. M. (2013). Sexually Dimorphic Neurons in the Ventromedial Hypothalamus Govern Mating in Both Sexes and Aggression in Males. *Cell*, 153(4), 896–909. <https://doi.org/10.1016/j.cell.2013.04.017>
- Young, H., Belbut, B., Baeta, M., & Petreanu, L. (2021). Laminar-specific cortico-cortical loops in mouse visual cortex. *ELife*, 10, 1–25. <https://doi.org/10.7554/eLife.59551>
- Zhou, L., Liu, M.-Z., Li, Q., Deng, J., Mu, D., & Sun, Y.-G. (2017). Organization of Functional Long-Range Circuits Controlling the Activity of Serotonergic Neurons in the Dorsal Raphe Nucleus. *Cell Reports*, 18(12), 3018–3032. <https://doi.org/10.1016/j.celrep.2017.02.077>
- Zuo, X.-N., Kelly, C., Di Martino, A., Mennes, M., Margulies, D. S., Bangaru, S., Grzadzinski, R., Evans, A. C., Zang, Y.-F., Castellanos, F. X., & Milham, M. P. (2010). Growing Together and Growing Apart: Regional and Sex Differences in the Lifespan Developmental Trajectories of Functional Homotopy. *Journal of Neuroscience*, 30(45), 15034–15043. <https://doi.org/10.1523/JNEUROSCI.2612-10.2010>
- Zuo, X.-N., He, Y., Betzel, R. F., Colcombe, S., Sporns, O., & Milham, M. P. (2017). Human Connectomics across the Life Span. *Trends in Cognitive Sciences*, 21(1), 32–45. <https://doi.org/10.1016/j.tics.2016.10.005>

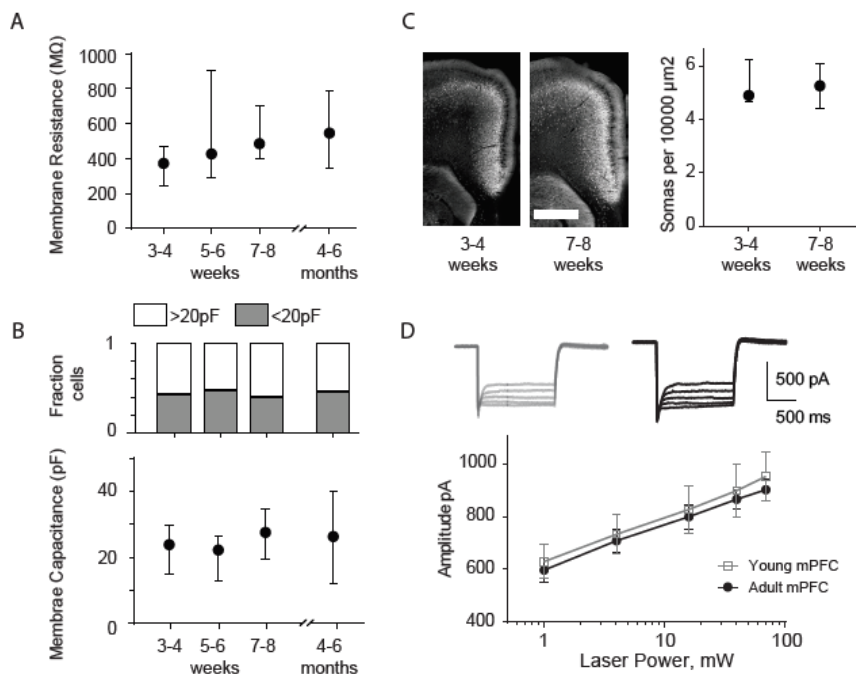
Author contributions

N.G.C., D.S., B.G. and Z.F.M. designed the experiments and analyses. N.G.C., D.S. and B.G. conducted the experiments. N.G.C. and D.S. analyzed the data. N.G.C. and D.S. wrote the original draft. Z.F.M. reviewed the manuscript.

Supplementaries

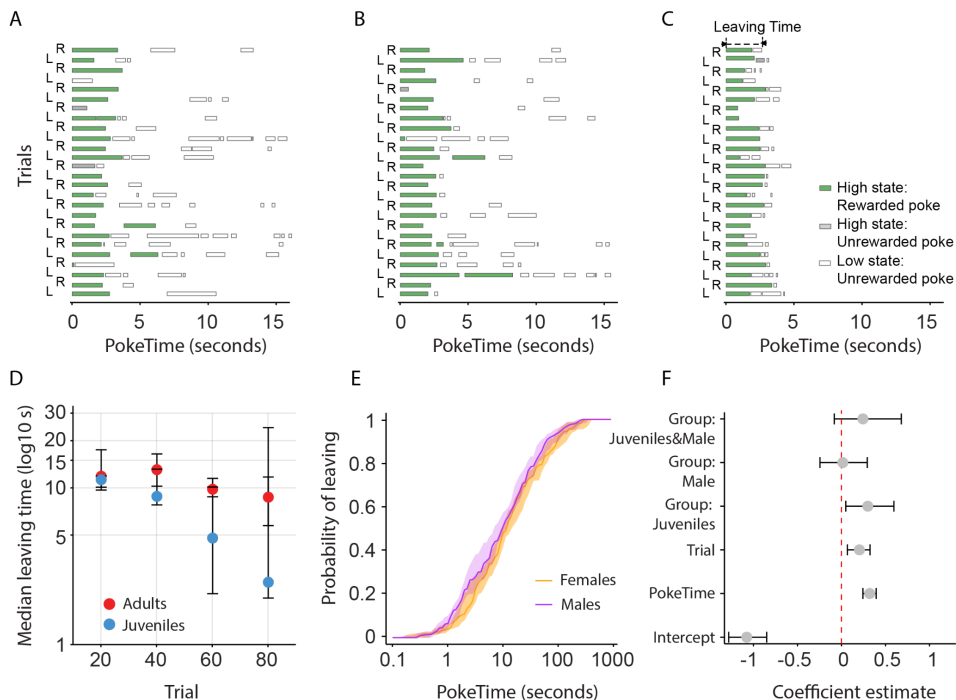


Supplementary figure 1. Changes in cortico-raphe connectivity over development are not explained by changes in the location of the recorded DRN neurons. (A) Example low magnification picture taken of a recorded DRN neuron and overlaid atlas inset used to determine its location. (B,C) Summary of the spatial location and connectivity of the recorded DRN neurons in juvenile (anterior DRN connected/non-connected = 1/16, posterior DRN connected/non-connected = 0/12) and adult (anterior DRN connected/non-connected=23/4, posterior DRN connected/non-connected = 13/4) mice.



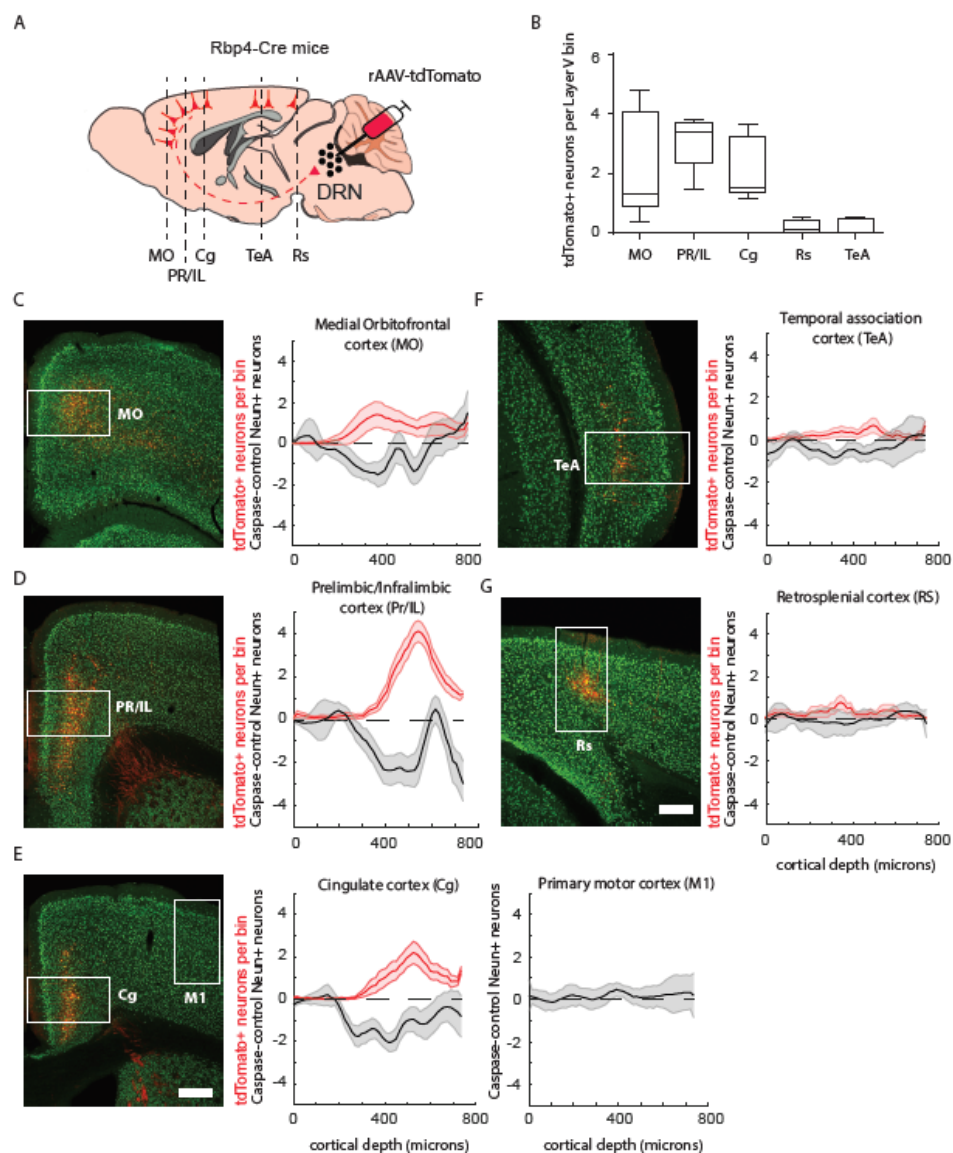
Supplementary figure 2. Changes in cortico-raphe connectivity over development are not explained by changes in membrane properties of DRN neurons or by differential ChR2 expression of ChR2 under the Rbp4 promoter over time. (A) The membrane resistance of DRN neurons is comparable over time (3-4 weeks: median = 400.4, 95% CI = [243.3, 500], 5-6 weeks: median = 427.4, 95% CI = [294.1, 925.2], 7-8 weeks: median = 485.4, 95% CI = [409.8, 699.3], 5-6 months: median = 546.4, 95% CI = [348.4, 806.5]). (B) The fraction of putative 5HT neurons (Capacitance>20pF) and non-5HT neurons (Capacitance<20pF) (Soiza-Reilly et al., 2019) recorded is comparable across developmental stages (Fraction of neurons with Capacitance>20pF: 3-4 weeks= 0.58, 5-6 weeks= 0.52, 7-8 weeks= 0.6, 5-6 months= 0.53. Chi-Square test χ^2 (3, N=116 neurons)=0.48, $p=0.92$). In addition, no overall changes in membrane capacitance were observed in DRN neurons across development (3-4 weeks: median = 23.8, 95% CI = [15, 29.5], 5-6 weeks: median =22.2, 95% CI = [12.7, 26.4], 7-8 weeks: median = 27.5, 95% CI = [19.3, 34.5], 5-6 months: median =26.2, 95% CI = [11.8, 40.1]). (C) The density of fluorescent mPFC layer V neurons is comparable in juvenile and adult Rbp4-tdTomato mice (3-4 weeks: median = 4.8, 95% CI = [4.5, 6.2], 7-8 weeks: median = 5.3, 95% CI = [4.3, 6]). (D) The evoked photocurrent in mPFC layer V neurons of juvenile and adult Rbp4-ChR2

mice is virtually identical across a wide range of stimulation intensities. Error bars in D represent mean $\pm$ SEM. Scale bar in C = 800 microns.



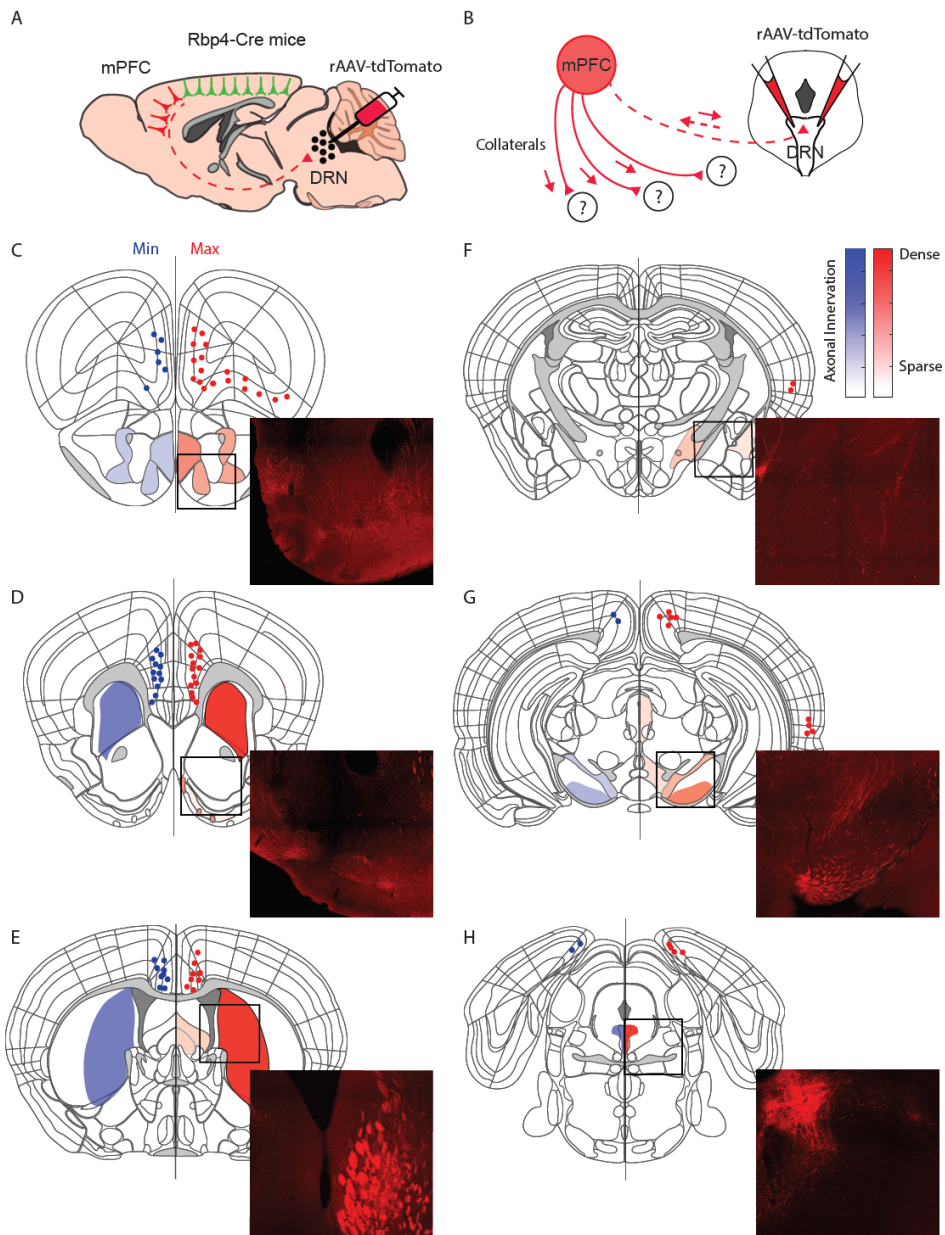
Supplementary figure 3. Behavior

(A-C) Examples of poking behavior over time for a naive adult (A), a naive juvenile (B), and a trained adult (C). Pokes in the active state can be rewarded (in green) or not (in gray). Pokes in the inactive state are never rewarded (in white). After the state switches, the mice have to travel to the other side (left or right port, L and R) to obtain more water. (D) Leaving time (median $\pm$ 95% CI across mice) as a function of elapsed trials in a session. (E) Cumulative distribution of the probability of leaving as a function of trial time elapsed (median $\pm$ 95% CI across mice) for female and male animals. (F) Regression coefficients $\pm$ 95% CI resulting from a parametric bootstrap ($n = 1000$) of a mixed models logistic regression to explain the probability of leaving. Note the lack of explanatory power for the group variable sex.



Supplementary figure 4. Labeling of Rbp4-expressing DRN projecting neurons with rAAV-tdTomato is consistent with cell density loss in mice injected with rAAV-Caspase3. (A) Schematic representation of rAAV-tdTomato dependent labeling of cortico-DRN projecting neurons in Rbp4-Cre mice. (B) Quantification of layer V tdTomato labeled somas across the different DRN projecting cortical areas (n = 8 mice). Example picture of the immunolabeling obtained with the pan-neuronal marker NeuN and the virally expressed tdTomato reporter together with the quantification across cortical depth of tdTomato cell

density and neuronal loss (NeuN density in rAAV-Caspase3 injected mice - average NeuN density of control mice) for the medial orbitofrontal cortex (C, Control vs. Caspase Two-sample Kolmogorov-Smirnoff Test, $D = 0.016$, $p = 0.09$), prelimbic/infralimbic cortex (D, $D = 0.026$, $p = 0.005$), cingulate cortex and motor primary cortex (E, $D = 0.022$, $p = 0.01$ and $D = 0.008$, $p = 0.29$, respectively), temporal association cortex (F, $D = 0.013$, $p = 0.14$) and retrosplenial cortex (G, $D = 0.015$, $p = 0.16$). Box plots represent median, IQR, and min/max data range. Shaded error plots represent mean $\pm$ SEM. Scale bar = 400 microns.



Supplementary figure 5. Dorsal raphe projecting cortical neurons have collateral projections that are particularly dense onto the mouse striatum. (A) Schematic representation of rAAV-tdTomato dependent labeling of cortico-DRN projecting neurons in Rbp4-Cre mice. (B) Schematic representation of axon collaterals from the same cortical subpopulation of neurons retrogradely labeled at the DRN. (C-H) Semiquantitative representation of axon collateral

innervation density across the anteroposterior levels of the mouse brain presenting the injections with the highest (red) and lowest (blue) density of retrogradely labeled neurons. Scale bar= 500 microns.

General discussion

This thesis's works span a diverse landscape of neuroscience interests: from inference to development and back to pharmacology. In the following section, we will discuss a circuit-level perspective to reunite the main findings collected along the way. To manage the topics' complexity, we will summarize the main conclusions across the various chapters. This reduction is functional to hypothesize how some of the different elements work in the circuit to modulate foraging decisions. However, it refers directly to the interpretation of our results that we find more likely to offer more explanatory power. Therefore, we refer the reader to each chapter's discussion to keep note of the caveats related to our interpretations. Next, we will propose a functional network comprising the investigated areas that could explain how the interactions in the circuits can modulate foraging. Finally, we will outline potential experimental approaches to test our circuit hypothesis.

Main findings

In brief, we can summarise our conclusions in 5 points that emerged from various results:

- Mice foraging in an artificial environment solve the patch departure problem consistently with the Marginal Value Theorem qualitative predictions (Chapter 2 and 3).
- Their behavior is better explained by adopting a heuristic comprising a reward reset probability of leaving and knowledge of the reward probability dynamics governing the environment states through inference (Chapter 2 and 4).

- Orbitofrontal cortex (OFC) activity integrity is crucial to infer uncued changes in the environment state governed by hidden variables (Chapter 4).
- Serotonin's (5-HT) levels can adjust the balance between persistence and flexibility during foraging. On the one hand, sustained 5-HT release, possibly through 5-HT_{2a} receptors' activation, promotes a more flexible behavior, increasing exploration. On the other hand, transient 5-HT release, conceivably acting on the 5-HT_{2c} receptors, enhances persistence, producing an increase in exploitation (Chapter 3).
- The maturation of connections between the medial prefrontal cortex (mPFC) to the dorsal raphe nuclei (DRN), the primary source of 5-HT to the forebrain, adjust the balance of exploration and exploitation towards a more persistent phenotype in adults compared to juveniles (Chapter 5).

A hypothetical circuitry

In formulating a putative circuitry governing foraging decision, it is useful, to begin with how the functions we attributed to the areas studied are involved in foraging. In Chapter 2 (Executive functions and the prefrontal cortex), we discuss in detail how having access to a state representation of the environment is crucial to adaptive behavior. Crucially the ability to detect state changes can greatly facilitate the solution of the foraging dilemma. Once states are defined, the question can be resolved by determining which action (exploit/explore) is best to perform in any state (e.g., Killeen and Fetterman, 1988). From our literature review, it emerged clearly that OFC plays a fundamental role in enabling animals to adapt behavior to changes in the environment flexibly (e.g., Roberts, 2006; Kolling et al., 2012; Izquierdo et al., 2017).

In particular, our results in Chapter 4 indicate that OFC inhibition prevents animals from detecting state changes to efficiently guide the balance between exploration and exploitation, reverting them to the use of a suboptimal stimulus-bond strategy (Vertechi et al., 2020).

Thinking about circuitry, we can ask what sort of information OFC requires to form or use state representations? State transitions cannot be immediately identified without explicit sensory cues; instead, they have to be inferred evaluating the likelihood of the recent events' history to belong to a given state (the "belief state," Rao, 2010; Wilson et al. 2014). Prediction error signals are one possible way to implement this evaluation, as more prediction errors occur, the lower the likelihood that the current "belief state" is correct. In a simplified task design with two options, as our foraging tasks, this should promote exploring the alternative site. Interestingly previous results from our and other labs have shown that 5-HT activity signals prediction errors (Nakamura et al., 2008; Inaba et al., 2013; Cohen et al., 2015; Hayashi et al., 2015; Matias et al., 2017). If 5-HT release in the OFC were related to prediction errors, this would be an effective downstream location to balance persistence and flexibility in terms of the likelihood of changing "belief state." In this hypothetical circuitry, when facing spurious prediction error, 5-HT release in OFC would be transient and preferentially active on the 5-HT_{2c} receptors. In this way, 5-HT would promote persistence, as in reducing the estimated likelihood that a state change has occurred. However, as more and possibly consecutive prediction errors occur, it would result in a sustained release of 5-HT in OFC, which on the contrary could enhance the estimated likelihood of a state change, therefore flexibility, via the 5-HT_{2a} receptors.

Although this specific hypothesis regarding the 5-HT role in OFC has never been directly tested, previous studies appear to support its

conceptual validity indirectly. First, it is firmly established that selective 5-HT depletion in the OFC impairs reversal learning (see, e.g., Roberts, 2011). More specifically and in line with our interpretation of Chapter 3 results, Furr and colleagues (2012) have shown that acute Serotonin reuptake inhibitor (SSRI, Citalopram) injection promotes flexibility in reversal tasks after stress exposure, and this effect depends on 5-HT_{2a} receptors in OFC. Furthermore, Boulougouris and Robbins (2010) have shown that 5-HT_{2c} receptors in OFC promote persistence in reversal tasks.

In striking contrast, selective optogenetic stimulation of 5-HT DRN axons in the OFC during a waiting task results in increased persistence rather than flexibility (Miyazaki et al., 2020). Unfortunately, to our knowledge, no study has directly manipulated OFC's 5-HT_{2a} and 5-HT_{2c} receptors in steady tasks (tasks in which the rules of the environment do not change after the stimuli- or action-outcome contingencies are learned). Yet, all studies that we examined, in which these receptors are studied, when the manipulation produced an effect, confirms the results from reversal tasks indicating that 5-HT_{2a} and 5-HT_{2c} respectively promote flexibility and persistence.

In conclusion, we propose that OFC governs the decision between exploitation and exploration in foraging, defining the current belief state. In this context, we postulate that one primary function of 5-HT is to balance the tendency towards flexibility and persistence. Mechanistically, we hypothesize that this effect occurs in OFC where transient 5-HT level increases activate 5-HT_{2c} receptors, strengthening persistence, while a sustained increase in 5-HT activates the 5-HT_{2a} receptors, promoting flexibility.

Testing circuitry hypothesis

In this section, we propose two possible experiments to test our hypothesis directly and address the limitation of our previous findings comparing serotonin reuptake inhibitors (SSRI) and optogenetic to mimic sustained and transient release of serotonin (5-HT), respectively.

As discussed in Chapter 3, optogenetic stimulation of DRN 5-HT neuron and systemic injection of Citalopram would act on not fully overlapping serotonergic targets. We propose substituting SSRI with designer receptors exclusively activated by designer drugs (DREADD) chemogenetic activation to solve this issue. Since both approaches rely on virus vector delivery, using a mixed viral solution, we can target to a large extent the same neurons in one individual, thus enabling us to compare in the same individual, across different sessions, the effects of sustained and transient 5-HT release during foraging. The second advantage of this approach is the possibility of using retrograde viruses, which allow us to restrict the manipulation to 5-HT neurons projecting to the orbitofrontal cortex (OFC). Ren and colleagues (2018) show that this anatomical distinction is crucial to separate functional differences within the serotonergic DRN circuit. We postulate with this preparation to observe an increase in persistence and flexibility when the DRN-OFC projections are stimulated with optogenetic and chemogenetic, respectively. However, it has to be noted that functional and anatomical separations exist within OFC as well (see e.g., Walton et al., 2011), and those might require a more precise definition of the target area.

Suppose the previous experiment results would confirm our hypothesis that sustained and transient 5-HT release promotes flexibility and persistence, respectively. In that case, we could take advantage of having localized the downstream target area to test the causal role of

5-HT receptors. We could directly assess their interaction by using local injections of selective antagonists of 5-HT_{2a} and 5-HT_{2c} receptors in the OFC, combined with a prolonged or transient 5-HT release. On the one hand, if our hypothesis is correct, we expect to observe that 5-HT_{2a} blockage in OFC should impair or block the increase in flexibility produced by sustained 5-HT release. On the other hand, we would expect the same effect of 5-HT_{2c} blocking on persistence produced by transient 5-HT release.

One aspect we have not investigated so far is the relationship between DRN serotonergic neuron activity and 5-HT release levels. Clarifying this result would be highly important to relate our downstream view with the knowledge acquired on DRN activity. Directly related to our work is a recent article from Grossman and colleagues (2020) that investigated DRN serotonergic neurons electrophysiology in relation to expected and unexpected uncertainty. This question is tightly related to our hypothesis for two main reasons. As expressed in more detail in Chapter 3, detecting the distinction between these two sources of uncertainty is a very likely candidate mechanism to recognize state transitions, one primary function attributed to OFC. Furthermore, one way to distinguish between the two forms of uncertainty is to keep track of how many and continuous prediction errors occurred, a type of signal we know is linked to 5-HT. Grossman and colleagues' findings show that 5-HT neurons signal expected uncertainty with sustained activity throughout a foraging trial, while they have bursts of activity on top to signal unexpected uncertainty. How does this regulate the 5-HT downstream level in OFC? Thanks to the development of genetically encoded sensors for 5-HT, we could investigate this question by combining photometry recordings in OFC with retrograde voltage or calcium sensors recording in the DRN. The unexpected

uncertainty-related burst of activity may result in a sustained enhanced level of 5-HT in OFC, which in our hypothesis should coherently promote 5-HT_{2a} drive flexibility (changing strategy in the face of unexpected uncertainty). Alternatively, the mechanism governing the bias towards the higher affinity persistence-promoting 5-HT_{2c} and the lower affinity flexibility-promoting 5-HT_{2a} could be governed not by temporal dynamics differences but pure concentration levels. In that case, this approach would still provide us with insight into how 5-HT uncertainty signals modulate the balance between persistence and flexibility.

Final remarks

Adaptive behavior is a complex task that brings together deduction about the changes in the environment and fine-tuning of the action and strategies at disposal. During my Ph.D., I had the privilege of exploring this topic from diverse perspectives and the fortune to collaborate with highly talented, bright, and supportive colleagues. In writing this thesis, I tried to strive to present new ideas and an accurate report of our results. I hope to have succeeded in respecting the quality and precision of the research performed by my colleagues while stimulating the reader with the sharing of untested ideas. Having access to this kind of hypothetical and forward-looking discussion has undoubtedly been my Ph.D.'s greatest gift, and this thesis would not represent it correctly if it were missing.

In conclusion, we presented evidence that mice engaged in a probabilistic foraging task are capable of inferring evidence regarding the statistical properties of the environment. This ability crucially relies on the prefrontal cortex network and the orbitofrontal cortex in particular. Furthermore, we propose a novel role for serotonin, to be in charge of fine-tuning exploration and exploitation during foraging bidirectionally.

References

- Cohen, J. Y., Amoroso, M. W., & Uchida, N. (2015). Serotonergic neurons signal reward and punishment on multiple timescales. *ELife*, 2015(4), 1–25. <https://doi.org/10.7554/eLife.06346>
- Furr, A., Lapiz-Bluhm, M. D., & Morilak, D. A. (2012). 5-HT_{2A}-receptors in the orbitofrontal cortex facilitate reversal learning and contribute to the beneficial cognitive effects of chronic citalopram treatment in rats. *The International Journal of Neuropsychopharmacology*, 15(09), 1295. <https://doi.org/10.1017/S1461145711001441>
- Grossman, C. D., Bari, B. A., & Cohen, J. Y. (2020). Serotonin neurons modulate learning rate through uncertainty. *BioRxiv*, 1. <https://doi.org/10.1101/2020.10.24.353508>
- Hayashi, K., Nakao, K., & Nakamura, K. (2015). Appetitive and aversive information coding in the primate dorsal raphe nucleus. *Journal of Neuroscience*, 35(15), 6195–6208. <https://doi.org/10.1523/JNEUROSCI.2860-14.2015>
- Inaba, K., Mizuhiki, T., Setogawa, T., Toda, K., Richmond, B. J., & Shidara, M. (2013). Neurons in Monkey Dorsal Raphe Nucleus Code Beginning and Progress of Step-by-Step Schedule, Reward Expectation, and Amount of Reward Outcome in the Reward Schedule Task. *Journal of Neuroscience*, 33(8), 3477–3491. <https://doi.org/10.1523/JNEUROSCI.4388-12.2013>
- Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, 345(March), 12–26. <https://doi.org/10.1016/j.neuroscience.2016.03.021>
- Killeen, P. R., & Fetterman, J. G. (1988). A Behavioral Theory of Timing. *Psychological Review*, 95(2), 274–295. <https://doi.org/10.1037/0033-295X.95.2.274>
- Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science*, 336(6077), 95–98. <https://doi.org/10.1126/science.1216930>
- Matias, S., Lottem, E., Dugué, G. P., & Mainen, Z. F. (2017). Activity patterns of serotonin neurons underlying cognitive flexibility. *ELife*, 6, 1–24. <https://doi.org/10.7554/eLife.20552>
- Miyazaki, K., Miyazaki, K. W., Sivori, G., Yamanaka, A., Tanaka, K. F., & Doya, K. (2020). Serotonergic projections to the orbitofrontal and medial prefrontal cortices differentially modulate waiting for future rewards. *Science Advances*, 6(48), 1–15. <https://doi.org/10.1126/sciadv.abc7246>

- Nakamura, K., Matsumoto, M., & Hikosaka, O. (2008). Reward-Dependent Modulation of Neuronal Activity in the Primate Dorsal Raphe Nucleus. *Journal of Neuroscience*, 28(20), 5331–5343.
<https://doi.org/10.1523/JNEUROSCI.0021-08.2008>
- Rao, R. P. N. (2010). Decision Making Under Uncertainty: A Neural Model Based on Partially Observable Markov Decision Processes. *Frontiers in Computational Neuroscience*, 4. <https://doi.org/10.3389/fncom.2010.00146>
- Ren, J., Friedmann, D., Xiong, J., Liu, C. D., Ferguson, B. R., Weerakkody, T., DeLoach, K. E., Ran, C., Pun, A., Sun, Y., Weissbourd, B., Neve, R. L., Huguenard, J., Horowitz, M. A., & Luo, L. (2018). Anatomically Defined and Functionally Distinct Dorsal Raphe Serotonin Sub-systems. *Cell*, 175(2), 472–487.e20. <https://doi.org/10.1016/j.cell.2018.07.043>
- Roberts, A. C. (2006). Primate orbitofrontal cortex and adaptive behaviour. *Trends in Cognitive Sciences*, 10(2), 83–90.
<https://doi.org/10.1016/j.tics.2005.12.002>
- Roberts, A. C. (2011). The importance of serotonin for orbitofrontal function. *Biological Psychiatry*, 69(12), 1185–1191.
<https://doi.org/10.1016/j.biopsych.2010.12.037>
- Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., Oude Lohuis, M. N., & Mainen, Z. F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. *Neuron*, 106(1), 166–176.e6. <https://doi.org/10.1016/j.neuron.2020.01.017>
- Walton, M. E., Behrens, T. E. J., Noonan, M. P., & Rushworth, M. F. S. (2011). Giving credit where credit is due: orbitofrontal cortex and valuation in an uncertain world. *Annals of the New York Academy of Sciences*, 1239(1), 14–24.
<https://doi.org/10.1111/J.1749-6632.2011.06257.X>
- Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, 81(2), 267–279.
<https://doi.org/10.1016/j.neuron.2013.11.005>

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

www.itqb.unl.pt