UNIVERSITY OF CALIFORNIA SAN DIEGO

Neither animals nor decisions are interchangeable; subjective experience shapes the brain and behavior

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Experimental Psychology

by

Drew Schreiner

Committee in charge:

Professor Christina Gremel, Chair
Professor Adam Aron
Professor Timothy Gentner
Professor Takaki Komiyama
Professor Byungkook Lim

2021

The dissertation of Drew Schreiner is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

I would like to dedicate this dissertation to my partner Meredith Chedsey. She has been a constant source of love, support, and encouragement throughout the long, sometimes arduous, process of graduate school. I am a better person for having her in my life.

EPIGRAPH

If one wishes to understand the behavior of animals, one must take account of their individuality, annoying as this may be for those who prefer the tidiness of physics, chemistry, and mathematical formulations.

Donald Griffin

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

1. Supplemental Tables: Tables 2.S1-2.S7 and Table 3.S1.

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge the indispensable support of Christina Gremel as my advisor and chair of my thesis committee. She has been an inspiration, and given me the training to excel and the space to pursue my research ambitions. I cannot imagine a better mentor in graduate school. She always seemed to know how best to help me learn and grow. I always knew that she genuinely cared and had my interests at heart. I would not be the scientist that I am today without her.

I would also like to acknowledge the support of everyone in the Gremel Lab, who I have constantly turned to for support and advice throughout my time in graduate school. There is no more helpful, thoughtful, intelligent, and just plain friendly collection of scientists. You all made life in the lab a daily joy.

I would also like to acknowledge the support and advice of my thesis committee, who have helped to improve this dissertation and inspired me as a scientist.

I would also like to acknowledge my friends and family. They have been an unshakeable source of support, and a welcome reminder of all the wonderful, exciting, worthwhile things there are in the world – only some of which lay in the realm of science.

Chapter 1 is, in full, a reprint of the material published in *Scientific Reports* (2018). Schreiner, Drew C; Gremel, Christina M, 2018. The dissertation author was the primary investigator and author of this paper. This project was funded by the NIH:

4R00AA021780-02 (C.M.G), AA026077-01A1 (C.M.G.) and Whitehall Foundation (C.M.G).

Chapter 2 is, in full, a reprint of the material submitted to *Nature Neuroscience*. Schreiner, Drew C.; Cazares, Christian; Renteria, Rafael; Gremel, Christina M. The dissertation author was the primary investigator and author of this paper. This project was funded by the NIH: F31AA027439 (D.C.S.), F32AA026776 (R.R.), R01AA026077 (C.M.G.), the NSF: DGE-1650112 (C.C.), and a Whitehall Foundation Award (C.M.G).

Chapter 3 is, in full, a reprint of the material to be submitted to *Biological Psychiatry*. Schreiner, Drew C.; Renteria, Rafael; Baltz, Emily T.; Cazares, Christian; Gremel, Christina M. The dissertation author was the primary investigator and author of this paper. This project was funded by the NIH: F31AA027439 (D.C.S.), F32AA026776 (R.R.), R01AA026077 (C.M.G.), the NSF: DGE-1650112 (C.C.), DGE-2038238 (E.T.B), and a Whitehall Foundation Award (C.M.G).

VITA

EDUCATION

University of Colorado at Boulder 2009-2013
B.A. Molecular, Cellular, and Developmental Biology
B.A. Psychology with focus in Neuroscience, *magna cum laude*

University of California San Diego 2015-2021
Ph.D. in Experimental Psychology
- Cognitive and Behavioral Neuroscience Specialization

RESEARCH EXPERIENCE

Graduate Student Researcher 2015-2021
University of California San Diego. Advisor: Dr. Christina Gremel

Laboratory Technician 2015
ScienceCare, Aurora Colorado

Professional Research Assistant 2013-2014
University of Colorado at Boulder. Advisor: Dr. Ryan Bachtell

Undergraduate Research Assistant 2011-2013
University of Colorado at Boulder. Advisor: Dr. Ryan Bachtell

RESEARCH GRANTS AND AWARDS

Enoch Gordis Award Finalist 2021
Research Society on Alcoholism Student Merit Award 2021
Cosyne New Attendees Travel Grant 2020
UC San Diego Graduate Student Administration Travel Grant 2020
Ruth L. Kirschstein National Research Service Award (F31AA027439) 2019-2021
Research Society on Alcoholism Student Merit Award 2019
Society for Neuroscience Trainee Professional Development Award 2018
Society for Neuroscience Chapter Grant Award (3 separate awards, for NeuwriteSD.org) 2018-2020
Research Society on Alcoholism Student Merit Award 2017
Magna cum laude Latin honors in Psychology 2013
Individual Howard Hughes Medical Institute Biomedical Sciences Grant 2012-2013
Undergraduate Research Opportunity Program Individual Grant 2012
Undergraduate Research Opportunity Program Assistantship Grant 2011-2012
Multiple Dean's Lists 2009-2013

PREPRINTS, IN PREPARATION, AND IN REVIEW

Schreiner, D.C., Cazares, C., Renteria, R., Gremel, C.M. (2021). Mice are not automatons; subjective experience in premotor circuits guides behavior. *BioRxiv*. https://doi.org/10.1101/2021.06.23.449617

Schreiner, D.C., Renteria, R., Baltz, E.T., Cazares, C., Gremel, C.M. (2021). Chronic alcohol exposure disrupts behavioral flexibility via hyperactive premotor corticostriatal activity. In Preparation.

PUBLICATIONS

Schreiner, D.C., Yalcinbas, E.A., Gremel, C.M. (2021). A push for examining subjective experience in value-based decision-making. *Current Opinion in Behavioral Sciences*, 41, 45-49. https://doi.org/10.1016/j.cobeha.2021.03.020

Cazares, C., **Schreiner, D.C.**, Gremel, C.M. (2021). Different effects of alcohol exposure on action and outcome related orbitofrontal cortex activity. *eNeuro*. https://doi.org/10.1523/ENEURO.0052-21.2021

Renteria, R., Cazares, C., Baltz, E.T., **Schreiner, D.C.**, Yalcinbas, E.A., Steinkellner, T., Hnasko, T.S., Gremel, C.M. (2021). Mechanisms for differential recruitment of orbitostriatal transmission during actions and outcomes following chronic alcohol exposure. *ELife*, 10, e67065. https://doi.org/10.7554/eLife.67065

Schreiner, D.C., Renteria, R., Gremel, C.M. (2019). Fractionating the all or none definition of goal-directed and habitual decision-making. *Journal of Neuroscience Research*, 98(6), 998-1006. https://doi.org/10.1002/jnr.24545

Schreiner, D.C., & Gremel, C.M. (2018). Orbital frontal cortex projections to secondary motor cortex mediate exploitation of learned rules. *Scientific Reports*, 8(1). https://doi.org/10.1038/s41598-018-29285-x

*Kavanagh, K. A., ***Schreiner, D.C.**, Levis, S. C., O'Neill, C. E., & Bachtell, R. K. (2015). Role of adenosine receptor subtypes in methamphetamine reward and reinforcement. *Neuropharmacology*, 89, 265–273. https://doi.org/10.1016/j.neuropharm.2014.09.030
*Equal Contribution

O'Neill, C. E., Levis, S. C., **Schreiner, D. C.**, Amat, J., Maier, S. F., & Bachtell, R. K. (2015). Effects of Adolescent Caffeine Consumption on Cocaine Sensitivity. *Neuropsychopharmacology*, 40(4), 813–821. https://doi.org/10.1038/npp.2014.278

2.1ABSTRACT OF THE DISSERTATION

Neither animals nor decisions are interchangeable; subjective experience shapes the brain and behavior

by

Drew Schreiner

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2021

Professor Christina Gremel, Chair

Decisions are not made in isolation. Rather, they rely on internal states, contextual, temporal, and historical information. This subjective experience is a powerful driver of behavior and the associated neural mechanisms. However, most neurobiological investigations of decision-making ignore the impact of subjective

experience and instead focus on constraining tasks (cues, trials) to isolate specific variables (choice, accuracy). By ignoring subjective experience and averaging across subjects and decisions, we may be left with an incomplete or inaccurate picture of the brain-behavior relationship. In this dissertation, I took a dual-pronged approach using relatively unconstrained tasks to investigate how subjective experience affects both behavior and the brain, with a focus on rodent premotor cortex (M2), as prior work has suggested that M2 is poised to be sensitive to this information. Chapter 1 explored how individual variation during learning affected decisions about when to explore versus exploit. Results suggested that individual experience with a rule strengthened exploitation of that rule, with projections from orbital frontal cortex to M2 necessary for this experience-based exploitation. Chapter 2 investigated what aspects of subjective experience were used to guide a self-paced, self-generated behavior. Mice used diverse sources of information beyond just prior actions and reward, including the passage of time and information-checking to guide decision-making. M2 integrated these information sources to bias strategy-level decision-making, while its projections into dorsal medial striatum (M2-DMS) were specifically necessary to implement a recent experience-based strategy. Chapter 3 explored how premotor function and sensitivity to subjective experience were affected by psychiatric disease. Prior chronic alcohol impaired behavioral flexibility, and this was causally linked to the induction of

hyperactivity of M2-DMS neurons, suggesting human premotor regions as novel therapeutic targets for alcohol use disorder. These studies show that diverse aspects of subjective experience powerfully drive behavior and its neural representation. They implicate premotor circuits in integrating subjective experience to drive flexible behavior, a role which may be disrupted in psychiatric disease. This suggests that attempts to ignore or factor out subjective experience may be misguided; whether we take account of it or not, it likely will affect the brain and behavior.

INTRODUCTION

Subjective experience is a powerful driver of behavior. By subjective experience, I mean the historical, temporal, contextual, consequential, associative, ethological, and internal factors continuously experienced by individual animals (Schreiner et al., 2021), all of which can affect decision-making (Ariely & Zakay, 2001; Balleine, 2019; Balleine & Dickinson, 1998; Berridge et al., 2008; Bouton & Balleine, 2019; Costa, 2011). This is distinct from the study of how individual differences or genetic variation ramify into different behavioral phenotypes; even genetically identical animals exposed to identical environments show variable behavior due to differing interactions with, and subjective experience of, that environment (Freund et al., 2013). Animals are not interchangeable automatons. But more than that, even superficially similar decisions made by the *same individual* may, in actuality, reflect different behavioral processes. For instance, either goal-directed or habitual learning systems are sufficient to acquire and perform a variety of behaviors and only careful investigation can unmask these behavioral controllers (Balleine, 2019). Thus, neither every animal, nor every decision is the same.

Yet all too often, neurobiological investigations into decision-making treat animals and decisions as interchangeable, ignoring how individual subjective

experience may have led to different behavioral control and neural representation, modulation, and recruitment. I would suggest that this is analogous to how research *used to be* conducted in paleontology and archaeology. Whereas older studies focused solely on the artifact or fossil (often destroying its surroundings in the process), modern approaches recognize that the context in which an artifact is embedded provides crucial information. What information are we missing by ignoring how subjective experience contributes to the brain and behavior? In each chapter of this dissertation, I try to address this question using a dual-pronged approach, investigating both how subjective experience affects behavior and how it may be controlled by the brain.

In contrast to most neurobiological investigations, natural decision-making is messy. It occurs continuously across time, as one moment bleeds into the next in a richly variable, non-stationary, open world. For instance, crows have the remarkable ability to fly up and drop shelled prey items (nuts, shellfish) to break them open. When they do this, they keep track of a staggering array of experiential variables, including, the hardness of the ground, the type of prey, the amount of kleptoparasitism, and even the number of times they dropped a specific item (Cristol & Switzer, 1999). While crows are relatively cognitively complex, use of subjective experience is widespread.

Like humans building cairns, wood mice leave distinct items as way-markers when exploring to aid navigation (Stopka & Macdonald, 2003). Even *drosophila* show experience-dependent differences in behavior and its neural representation (Jacob et al., 2021).

Indeed, the drive to use experience to guide behavior is so strong that humans and other animals use their experience even when it is unnecessary or actively harmful. In most perceptual decision-making tasks subjects “should” pay attention only to the current sensory stimulus – but they do not. Instead, these tasks show strong history dependencies (e.g., Busse et al., 2011). By leveraging these dependencies, careful investigation has revealed that there is a principled mechanism involved in how mice, rats, and humans use their experience. Namely, reward biases (e.g., win-stay/lost-shift) are particularly strong when decision confidence is low (Lak et al., 2020), demonstrating that directed study of experience can provide novel insight.

The preponderance of non-stationary environments in the natural world and this propensity to use subjective experience even when it is “unnecessary” suggests that animals are well-adapted to use experience to guide their behavior. The brain should therefore represent and control the use of this experiential information. However, neurobiological studies typically interpret neural activity in relation to a snapshot of

behavior using discrete, averaged task-related variables (e.g. total actions, accuracy, etc.). Even in studies directly interrogating experience (e.g., Botta et al., 2020; Dhawale et al., 2019; Hattori et al., 2019; Hwang et al., 2019; Iigaya et al., 2018; Pisupati et al., 2021), tasks are often constrained (discrete variables, binary choice, trials, cues, restraint, etc.) which, aside from being rather different from the natural world (analog variables, many choices, continuous time, inconsistent/evolving cues, freely moving, etc.), may impair investigation of subjective experience. For instance, trials make it difficult to assess how decision-making evolves continuously across time, and cues may lead to elicited behavior, which can recruit different neural circuits relative to voluntary or self-generated behavior (Yin & Knowlton, 2006). In this dissertation, I use un-cued, self-generated behavior with analog variables to address these potential issues and investigate several outstanding questions about subjective experience including: how do different learning environments and individual variation within an environment affect *what* is learned (Chapter 1), what aspects of subjective experience contribute to decision-making (Chapters 2-3), and how is sensitivity to experience altered in psychiatric disease (Chapter 3).

While it is understandable to limit and constrain investigations to increase experimental control, there is a real risk that the mechanisms identified by these

approaches provide an incomplete or inaccurate picture, and this may be particularly relevant for psychiatric disorders which, of course, occur in a fully open world. Disorders such as Obsessive Compulsive Disorder (OCD) or Substance Use Disorders are in part characterized by disrupted decision-making, including habitual or compulsive decisions (Everitt & Robbins, 2005, 2016; Gillan et al., 2011; Graybiel & Rauch, 2000). Put another way - actions that are inappropriately sensitive to experience and feedback (e.g., insensitivity to the negative consequences of actions in compulsivity). These psychiatric disorders are associated with disruption to prefrontal and premotor cortex, as well as corticostriatal circuits (e.g. for review, Gillan & Robbins, 2014; Graybiel & Rauch, 2000; Gremel & Lovinger, 2017; Lovinger & Gremel, 2021). Importantly, unlike deeper brain structures that are the focus of many circuit-based investigations, premotor cortices in humans are dorsally located and accessible to treatments such as transcranial magnetic stimulation (TMS), which have been reported to reduce compulsivity in OCD (Gomes et al., 2012; Hawken et al., 2016; Mantovani et al., 2013). Thus, in this dissertation I will focus on the role of premotor circuits in subjective experience, as mechanistic insight here has the potential not only to advance our understanding of the brain-behavior relationship in relation to subjective experience, but may also improve the development and targeting of novel treatment strategies.

Aside from this potential treatment relevance, premotor circuits are also poised to be sensitive to and use experience to guide appropriately flexible behavior. In humans, this includes pre-supplementary/supplementary motor areas (Pre-SMA/SMA), while the rodent homologue is the premotor cortex (M2, also referred to herein as secondary motor cortex) (Barthas & Kwan, 2017). M2 receives input from a variety of sensory and associative cortical and thalamic regions (Reep et al., 1987, 1990; Zingg et al., 2014), is reciprocally connected with M1 (Reep et al., 1987), and has extensive projections into basal ganglia regions, particularly the dorsal striatum (Delevich et al., 2020; Hintiryan et al., 2016). Existing research has suggested that M2 is involved in choice (Barthas & Kwan, 2017; Steinmetz et al., 2019; Zatka-Haas et al., 2021), evidence accumulation (Erlich et al., 2011, 2015; Hanks et al., 2015; Orsolic et al., 2021; Pinto et al., 2019), skilled motor learning (Cao et al., 2015; Makino et al., 2017), and using history (e.g., prior actions and outcomes) to guide decision-making (Hattori et al., 2019; Pisupati et al., 2021; Siniscalchi et al., 2016, 2019; Sul et al., 2011). It thus seems evident that M2 is sensitive to history. However, it is unclear how M2 may be sensitive to subjective experience, particularly in an un-cued, self-paced, self-generated context. There remain many open questions about how M2 receives, represents, and uses subjective experience. These questions include: what information is M2 receiving from associative cortical regions (Chapter 1), what types of

subjective experience does M2 represent and use to guide decision-making (Chapter 2), what information does M2 convey to downstream regions such as the dorsal striatum (Chapters 2-3), and how does psychiatric disease affect M2 activity and function (Chapter 3).

First, addressing the question of M2 input; Orbitofrontal cortex (OFC) is one associative region that sends projections into M2 (Zingg et al., 2014), and correlative studies of OFC-M2 projections have shown that plasticity of this projection is associated with rule learning (Johnson et al., 2016). However, it is unclear if OFC-M2 input is causally necessary for the exploitation of known rules. Behaviorally, while it is known that different types of reinforcement schedules can bias different decision strategies (Adams, 1982; Adams & Dickinson, 1981; Derusso et al., 2010; Dickinson & Balleine, 1994; Hilario et al., 2007, 2012), it is unclear how individual variation *within* identical schedules may contribute, nor is it clear how individual variation affects decisions about when to exploit or explore. I set out to address these questions in Chapter 1 by training mice to press one lever for a food reward before introducing a novel lever at test to determine how subjective experience influenced decisions about exploiting the trained lever versus exploring the novel lever. Results suggested that the degree of experience individual mice had with a rule strongly predicted subsequent

exploitation of that rule. Chemogenetic inhibition of OFC-M2 projections during learning reduced exploitation of the learned rule, suggesting a causal link between OFC-M2 activity/plasticity and rule learning.

Although Chapter 1 showed that individual variation during learning can affect strategy-level decisions, it did not provide insight into what aspects of subjective experience mice used to control their decision-making, or how these aspects might be represented in M2. In Chapter 2, I investigated these questions. Mice were trained to press and hold down a lever for at least a minimum duration to earn a food reward (Fan et al., 2012; Platt et al., 1973; Yin, 2009). There were no trials or cues and reward was delivered at offset. Thus, in this self-paced, self-generated task experience was essential for performance, allowing for the investigation and modeling of how animals used subjective experience to guide decision-making. Results suggested that mice learned to rely on both recent and long-term subjective experience, and that typically-neglected variables such as elapsed time and information checking behavior played a large role in guiding flexible behavior. Contrary to prevailing dogma from binary choice tasks, reward played little role in determining subsequent behavior. Following up on the examination of OFC input into M2, Chapter 2 turned towards examining M2 itself and its Dorsal Medial Striatum projections (M2-DMS). M2

integrated various sources of subjective experience and used this to bias strategy level decisions. M2-DMS projections used specific aspects of recent experience to plan upcoming actions, implicating M2-DMS in enacting an experience-based strategy. These results show that neglected aspects of experience can potently affect both behavior and associated neural activity, suggesting that such information should not be ignored or factored out as irrelevant (Roy et al., 2021).

These same premotor corticostriatal circuits are disrupted in psychiatric diseases such as Alcohol Use Disorder (AUD). AUD is associated with altered premotor function as well as impaired behavioral flexibility in humans (Claus et al., 2011; Duka et al., 2011; Sjoerds et al., 2014). While these correlative studies are suggestive, there has been no demonstration of a causal link. Additionally, it is unclear precisely how AUD affects behavioral flexibility as a number of different computations support flexible behavior (Schreiner et al., 2020; Shnitko et al., 2020). Nor is it clear how chronic alcohol might specifically affect the activity, recruitment, and function of M2, as alcohol effects on M2 have only been investigated in the context of brain-wide MRI or cFos studies (Dudek et al., 2015; Liu & Crews, 2015). In Chapter 3, I addressed these important gaps. Mice were repeatedly exposed to chronic intermittent ethanol vapor (or air control) and withdrawal using a well-validated rodent model of

alcohol dependence (Becker & Hale, 1993). During protracted withdrawal from chronic alcohol (Heilig et al., 2010), mice were trained in the lever hold down task described above. Results suggested that prior chronic alcohol specifically impaired the use of recent subjective experience in guiding flexible behavior. Furthermore, chronic alcohol caused long-lasting alterations to the intrinsic properties of M2 projection neurons, and specifically induced hyperactivity in M2-DMS projection neurons during decision-making. Finally, chemogenetic inhibition of this alcohol-induced hyperactive M2-DMS rescued appropriate behavioral flexibility, demonstrating a causal link between chronic alcohol's effects on behavioral flexibility and M2-DMS activity. Importantly, this was true during protracted withdrawal from chronic alcohol, and in a food-based foraging task, suggesting that this is a long-lasting neuroadaptation with generalized behavioral deficits. Furthermore, these results suggest human premotor regions as a therapeutic target in AUD.

Collectively, these three chapters fill an important void in neurobiological research. They demonstrate that individual subjective experience can play a large role in flexible behavior, with neglected aspects of this experience determining strategy-level decisions and circuit recruitment. Subjective experience was heavily represented in premotor circuits, both input into M2 from OFC (Chapter 1), within M2 itself (Chapter

2), and in the projection specific M2-DMS population (Chapters 2-3). By incorporating subjective experience into the investigation, results suggested a causal link between AUD disruption to premotor function and impaired behavioral flexibility (Chapter 3). Accounting for subjective experience has the potential to reveal fundamental insights into the brain and behavior, and these insights may also prove relevant for the treatment of psychiatric disease.

References

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *34*(2), 77–98. https://doi.org/10.1080/14640748208400878

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *33*(2), 109–121. https://doi.org/10.1080/14640748108400816

Ariely, D., & Zakay, D. (2001). A timely account of the role of duration in decision making. *Acta Psychologica*, *108*(2), 187–207. https://doi.org/10.1016/S0001-6918(01)00034-8

Balleine, B. W. (2019). The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. *Neuron*, *104*(1), 47–62. https://doi.org/10.1016/j.neuron.2019.09.024

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4–5), 407–419. https://doi.org/10.1016/S0028-3908(98)00033-1

Barthas, F., & Kwan, A. C. (2017). Secondary Motor Cortex: Where 'Sensory' Meets 'Motor' in the Rodent Frontal Cortex. *Trends in Neurosciences*, *40*(3), 181–193. https://doi.org/10.1016/j.tins.2016.11.006

Becker, H. C., & Hale, R. L. (1993). Repeated Episodes of Ethanol Withdrawal Potentiate the Severity of Subsequent Withdrawal Seizures: An Animal Model of Alcohol Withdrawal "Kindling." *Alcoholism: Clinical and Experimental Research*, *17*(1), 94–98. https://doi.org/10.1111/j.1530-0277.1993.tb00731.x

Berridge, K. C., Zhang, J., & Aldridge, J. W. (2008). Computing motivation: Incentive salience boosts of drug or appetite states. *The Behavioral and Brain Sciences*, *31*(4), 440–441. https://doi.org/10.1017/S0140525X08004767

Botta, P., Fushiki, A., Vicente, A. M., Hammond, L. A., Mosberger, A. C., Gerfen, C. R., Peterka, D., & Costa, R. M. (2020). An Amygdala Circuit Mediates Experience-Dependent Momentary Arrests during Exploration. *Cell*, *183*(3), 605-619.e22. https://doi.org/10.1016/j.cell.2020.09.023

Bouton, M. E., & Balleine, B. W. (2019). Prediction and control of operant behavior: What you see is not all there is. *Behavior Analysis: Research and Practice*, *19*(2), 202–212. https://doi.org/10.1037/bar0000108

Busse, L., Ayaz, A., Dhruv, N. T., Katzner, S., Saleem, A. B., Schölvinck, M. L., Zaharia, A. D., & Carandini, M. (2011). The Detection of Visual Contrast in the

Behaving Mouse. *The Journal of Neuroscience*, *31*(31), 11351–11361. https://doi.org/10.1523/JNEUROSCI.6689-10.2011

Cao, V. Y., Ye, Y., Mastwal, S., Ren, M., Coon, M., Liu, Q., Costa, R. M., & Wang, K. H. (2015). Motor Learning Consolidates Arc-Expressing Neuronal Ensembles in Secondary Motor Cortex. *Neuron*, *86*(6), 1385–1392. https://doi.org/10.1016/j.neuron.2015.05.022

Claus, E. D., Kiehl, K. A., & Hutchison, K. E. (2011). Neural and Behavioral Mechanisms of Impulsive Choice in Alcohol Use Disorder. *Alcoholism, Clinical and Experimental Research*, *35*(7), 1209–1219. https://doi.org/10.1111/j.1530-0277.2011.01455.x

Costa, R. M. (2011). A selectionist account of de novo action learning. *Current Opinion in Neurobiology*, *21*(4), 579–586. https://doi.org/10.1016/j.conb.2011.05.004

Cristol, D. A., & Switzer, P. V. (1999). Avian prey-dropping behavior. II. American crows and walnuts. *Behavioral Ecology*, *10*(3), 220–226. https://doi.org/10.1093/beheco/10.3.220

Delevich, K., Okada, N. J., Rahane, A., Zhang, Z., Hall, C. D., & Wilbrecht, L. (2020). Sex and Pubertal Status Influence Dendritic Spine Density on Frontal Corticostriatal Projection Neurons in Mice. *Cerebral Cortex*, *30*(6), 3543–3557. https://doi.org/10.1093/cercor/bhz325

Derusso, A. L., Fan, D., Gupta, J., Shelest, O., Costa, R. M., & Yin, H. H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. *Frontiers in Integrative Neuroscience*, *4*, 17. https://doi.org/10.3389/fnint.2010.00017

Dhawale, A. K., Miyamoto, Y. R., Smith, M. A., & Ölveczky, B. P. (2019). Adaptive Regulation of Motor Variability. *Current Biology*, *29*(21), 3551-3562.e7. https://doi.org/10.1016/j.cub.2019.08.052

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1–18.

Dudek, M., Abo-Ramadan, U., Hermann, D., Brown, M., Canals, S., Sommer, W. H., & Hyytiä, P. (2015). Brain activation induced by voluntary alcohol and saccharin drinking in rats assessed with manganese-enhanced magnetic resonance imaging. *Addiction Biology*, *20*(6), 1012–1021. https://doi.org/10.1111/adb.12179

Duka, T., Trick, L., Nikolaou, K., Gray, M. A., Kempton, M. J., Williams, H., Williams, S. C. R., Critchley, H. D., & Stephens, D. N. (2011). Unique Brain Areas Associated with Abstinence Control Are Damaged in Multiply Detoxified

Alcoholics. *Biological Psychiatry*, *70*(6), 545–552. https://doi.org/10.1016/j.biopsych.2011.04.006

Erlich, J. C., Bialek, M., & Brody, C. D. (2011). A Cortical Substrate for Memory-Guided Orienting in the Rat. *Neuron*, *72*(2), 330–343. https://doi.org/10.1016/j.neuron.2011.07.010

Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., & Brody, C. D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *ELife*, *4*, e05457. https://doi.org/10.7554/eLife.05457

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, *8*(11), 1481–1489. https://doi.org/10.1038/nn1579

Everitt, B. J., & Robbins, T. W. (2016). Drug Addiction: Updating Actions to Habits to Compulsions Ten Years On. *Annual Review of Psychology*, *67*(1), 23–50. https://doi.org/10.1146/annurev-psych-122414-033457

Fan, D., Rossi, M. A., & Yin, H. H. (2012). Mechanisms of Action Selection and Timing in Substantia Nigra Neurons. *Journal of Neuroscience*, *32*(16), 5534–5548. https://doi.org/10.1523/JNEUROSCI.5924-11.2012

Freund, J., Brandmaier, A. M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., Sachser, N., Lindenberger, U., & Kempermann, G. (2013). Emergence of Individuality in Genetically Identical Mice. *Science*, *340*(6133), 756–759. https://doi.org/10.1126/science.1235294

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., & de Wit, S. (2011). Disruption in the Balance Between Goal-Directed Behavior and Habit Learning in Obsessive-Compulsive Disorder. *The American Journal of Psychiatry*, *168*(7), 718–726. https://doi.org/10.1176/appi.ajp.2011.10071062

Gillan, C. M., & Robbins, T. W. (2014). Goal-directed learning and obsessive–compulsive disorder. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655). https://doi.org/10.1098/rstb.2013.0475

Gomes, P. V. O., Brasil-Neto, J. P., Allam, N., & Rodrigues de Souza, E. (2012). A Randomized, Double-Blind Trial of Repetitive Transcranial Magnetic Stimulation in Obsessive-Compulsive Disorder With Three-Month Follow-Up. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *24*(4), 437–443. https://doi.org/10.1176/appi.neuropsych.11100242

Graybiel, A. M., & Rauch, S. L. (2000). Toward a Neurobiology of Obsessive-Compulsive Disorder. *Neuron*, *28*(2), 343–347. https://doi.org/10.1016/S0896-6273(00)00113-6

Gremel, C. M., & Lovinger, D. M. (2017). Associative and sensorimotor cortico-basal ganglia circuit roles in effects of abused drugs. *Genes, Brain and Behavior*, *16*(1), 71–85. https://doi.org/10.1111/gbb.12309

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, *520*(7546), 220–223. https://doi.org/10.1038/nature14066

Hattori, R., Danskin, B., Babic, Z., Mlynaryk, N., & Komiyama, T. (2019). Area-Specificity and Plasticity of History-Dependent Value Coding During Learning. *Cell*, *177*(7), 1858-1872.e15. https://doi.org/10.1016/j.cell.2019.04.027

Hawken, E. R., Dilkov, D., Kaludiev, E., Simek, S., Zhang, F., & Milev, R. (2016). Transcranial Magnetic Stimulation of the Supplementary Motor Area in the Treatment of Obsessive-Compulsive Disorder: A Multi-Site Study. *International Journal of Molecular Sciences*, *17*(3). https://doi.org/10.3390/ijms17030420

Heilig, M., Egli, M., Crabbe, J. C., & Becker, H. C. (2010). Acute withdrawal, protracted abstinence and negative affect in alcoholism: Are they linked? *Addiction Biology*, *15*(2), 169–184. https://doi.org/10.1111/j.1369-1600.2009.00194.x

Hilario, M., Clouse, E., Yin, H. H., & Costa, R. M. (2007). Endocannabinoid Signaling is Critical for Habit Formation. *Frontiers in Integrative Neuroscience*, *1*. https://doi.org/10.3389/neuro.07.006.2007

Hilario, M., Holloway, T., Jin, X., & Costa, R. M. (2012). Different dorsal striatum circuits mediate action discrimination and action generalization: Neural circuits underlying action generalization. *European Journal of Neuroscience*, *35*(7), 1105–1114. https://doi.org/10.1111/j.1460-9568.2012.08073.x

Hintiryan, H., Foster, N. N., Bowman, I., Bay, M., Song, M. Y., Gou, L., Yamashita, S., Bienkowski, M. S., Zingg, B., Zhu, M., Yang, X. W., Shih, J. C., Toga, A. W., & Dong, H.-W. (2016). The mouse cortico-striatal projectome. *Nature Neuroscience*, *19*(8), 1100–1114. https://doi.org/10.1038/nn.4332

Hwang, E. J., Link, T. D., Hu, Y. Y., Lu, S., Wang, E. H.-J., Lilascharoen, V., Aronson, S., O'Neil, K., Lim, B. K., & Komiyama, T. (2019). Corticostriatal Flow of Action Selection Bias. *Neuron*, *104*(6), 1126-1140.e6. https://doi.org/10.1016/j.neuron.2019.09.028

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), 1–10. https://doi.org/10.1038/s41467-018-04840-2

Jacob, P. F., Vargas-Gutierrez, P., Okray, Z., Vietti-Michelina, S., Felsenberg, J., & Waddell, S. (2021). Prior experience conditionally inhibits the expression of new learning in Drosophila. *Current Biology*, *31*(16), 3490-3503.e3. https://doi.org/10.1016/j.cub.2021.05.056

Johnson, C. M., Peckler, H., Tai, L.-H., & Wilbrecht, L. (2016). Rule learning enhances structural plasticity of long-range axons in frontal cortex. *Nature Communications*, *7*, 10785. https://doi.org/10.1038/ncomms10785

Lak, A., Hueske, E., Hirokawa, J., Masset, P., Ott, T., Urai, A. E., Donner, T. H., Carandini, M., Tonegawa, S., Uchida, N., & Kepecs, A. (2020). Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. *ELife*, *9*, e49834. https://doi.org/10.7554/eLife.49834

Liu, W., & Crews, F. T. (2015). Adolescent Intermittent ethanol exposure enhances ethanol activation of the nucleus accumbens while blunting the prefrontal cortex responses in adult rat. *Neuroscience*, *293*, 92–108. https://doi.org/10.1016/j.neuroscience.2015.02.014

Lovinger, D. M., & Gremel, C. M. (2021). A Circuit-Based Information Approach to Substance Abuse Research. *Trends in Neurosciences*, *44*(2), 122–135. https://doi.org/10.1016/j.tins.2020.10.005

Makino, H., Ren, C., Liu, H., Kim, A. N., Kondapaneni, N., Liu, X., Kuzum, D., & Komiyama, T. (2017). Transformation of Cortex-wide Emergent Properties during Motor Learning. *Neuron*, *94*(4), 880-890.e8. https://doi.org/10.1016/j.neuron.2017.04.015

Mantovani, A., Rossi, S., Bassi, B. D., Simpson, H. B., Fallon, B. A., & Lisanby, S. H. (2013). Modulation of motor cortex excitability in obsessive-compulsive disorder: An exploratory study on the relations of neurophysiology measures with clinical outcome. *Psychiatry Research*, *210*(3), 1026–1032. https://doi.org/10.1016/j.psychres.2013.08.054

Orsolic, I., Rio, M., Mrsic-Flogel, T. D., & Znamenskiy, P. (2021). Mesoscale cortical dynamics reflect the interaction of sensory evidence and temporal expectation during perceptual decision-making. *Neuron*, *109*(11), 1861-1875.e10. https://doi.org/10.1016/j.neuron.2021.03.031

Pinto, L., Rajan, K., DePasquale, B., Thiberge, S. Y., Tank, D. W., & Brody, C. D. (2019). Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions. *Neuron*, *104*(4), 810-824.e9. https://doi.org/10.1016/j.neuron.2019.08.025

Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., & Churchland, A. K. (2021). Lapses in perceptual decisions reflect exploration. *ELife*, *10*, e55490. https://doi.org/10.7554/eLife.55490

Platt, J. R., Kuch, D. O., & Bitgood, S. C. (1973). Rats' lever-press durations as psychophysical judgments of time. *Journal of the Experimental Analysis of Behavior*, *19*(2), 239–250. https://doi.org/10.1901/jeab.1973.19-239

Reep, R. L., Corwin, J. V., Hashimoto, A., & Watson, R. T. (1987). Efferent Connections of the Rostral Portion of Medial Agranular Cortex in Rats. *Brain Research Bulletin*, *19*(2), 203–221. https://doi.org/10.1016/0361-9230(87)90086-4

Reep, R. L., Goodwin, G. S., & Corwin, J. V. (1990). Topographic organization in the corticocortical connections of medial agranular cortex in rats. *The Journal of Comparative Neurology*, *294*(2), 262–280. https://doi.org/10.1002/cne.902940210

Roy, N. A., Bak, J. H., Akrami, A., Brody, C. D., & Pillow, J. W. (2021). Extracting the dynamics of behavior in sensory decision-making experiments. *Neuron*, *109*(4), 597-610.e6. https://doi.org/10.1016/j.neuron.2020.12.004

Schreiner, D. C., Renteria, R., & Gremel, C. M. (2020). Fractionating the all-or-nothing definition of goal-directed and habitual decision-making. *Journal of Neuroscience Research*, *98*(6), 998–1006. https://doi.org/10.1002/jnr.24545

Schreiner, D. C., Yalcinbas, E. A., & Gremel, C. M. (2021). A push for examining subjective experience in value-based decision-making. *Current Opinion in Behavioral Sciences*, *41*, 45–49. https://doi.org/10.1016/j.cobeha.2021.03.020

Shnitko, T. A., Gonzales, S. W., Newman, N., & Grant, K. A. (2020). Behavioral Flexibility in alcohol drinking monkeys: The morning after. *Alcoholism, Clinical and Experimental Research*, *44*(3), 729–737. https://doi.org/10.1111/acer.14289

Siniscalchi, M. J., Phoumthipphavong, V., Ali, F., Lozano, M., & Kwan, A. C. (2016). Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior. *Nature Neuroscience*, *19*(9), 1234–1242. https://doi.org/10.1038/nn.4342

Siniscalchi, M. J., Wang, H., & Kwan, A. C. (2019). Enhanced Population Coding for Rewarded Choices in the Medial Frontal Cortex of the Mouse. *Cerebral Cortex*, *29*(10), 4090–4106. https://doi.org/10.1093/cercor/bhy292

Sjoerds, Z., Brink, W. van den, Beekman, A. T. F., Penninx, B. W. J. H., & Veltman, D. J. (2014). Response inhibition in alcohol-dependent patients and patients with depression/anxiety: A functional magnetic resonance imaging study. *Psychological Medicine*, *44*(8), 1713–1725. https://doi.org/10.1017/S0033291713002274

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 1–8. https://doi.org/10.1038/s41586-019-1787-x

Stopka, P., & Macdonald, D. W. (2003). Way-marking behaviour: An aid to spatial navigation in the wood mouse (Apodemus sylvaticus). *BMC Ecology*, *3*(1), 3. https://doi.org/10.1186/1472-6785-3-3

Sul, J. H., Jo, S., Lee, D., & Jung, M. W. (2011). Role of rodent secondary motor cortex in value-based action selection. *Nature Neuroscience*, *14*(9), 1202–1208. https://doi.org/10.1038/nn.2881

Yin, H. H. (2009). The role of the murine motor cortex in action duration and order. *Frontiers in Integrative Neuroscience*, *3*, 23. https://doi.org/10.3389/neuro.07.023.2009

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476. https://doi.org/10.1038/nrn1919

Zatka-Haas, P., Steinmetz, N. A., Carandini, M., & Harris, K. D. (2021). Sensory coding and the causal impact of mouse cortex in a visual decision. *ELife*, *10*, e63163. https://doi.org/10.7554/eLife.63163

Zingg, B., Hintiryan, H., Gou, L., Song, M. Y., Bay, M., Bienkowski, M. S., Foster, N. N., Yamashita, S., Bowman, I., Toga, A. W., & Dong, H.-W. (2014). Neural Networks of the Mouse Neocortex. *Cell*, *156*(5), 1096–1111. https://doi.org/10.1016/j.cell.2014.02.023

CHAPTER ONE

Orbital Frontal Cortex Projections to Secondary Motor Cortex Mediate Exploitation of Learned Rules

Drew C. Schreiner[1] and Christina M. Gremel[1, 2]

[1]Department of Psychology, University of California San Diego, La Jolla, California, 92093

[2]Neurosciences Graduate Program, University of California, San Diego, La Jolla, California, 92093

Chapter One is a reprint of material that appeared in *Scientific Reports*, 8, 2018.

Abstract

Animals face the dilemma between exploiting known opportunities and exploring new ones, a decision-making process supported by cortical circuits. While different types of learning may bias exploration, the circumstances and the degree to which bias occurs is unclear. We used an instrumental lever press task in mice to examine whether learned rules generalize to exploratory situations and the cortical circuits involved. We first trained mice to press one lever for food and subsequently assessed how that learning influenced pressing of a second novel lever. Using outcome devaluation procedures we found that novel lever exploration was not dependent on the food value associated with the trained lever. Further, changes in the temporal uncertainty of when a lever press would produce food did not affect exploration. Instead, accrued experience with the instrumental contingency was strongly predictive of test lever pressing with a positive correlation between experience and trained lever exploitation, but not novel lever exploration. Chemogenetic attenuation of orbital frontal cortex (OFC) projection into secondary motor cortex (M2) biased novel lever exploration, suggesting that experience increases OFC-M2 dependent exploitation of learned associations but leaves exploration constant. Our data suggests exploitation and exploration are parallel decision-making systems that do not necessarily compete.

Introduction

The concepts of exploration and exploitation have been widely studied with focus on the competition between these two processes (Addicott et al., 2017; Cohen et al., 2007). However, the classical conception of this dilemma (Sutton & Barto, 1998) often neglects the possibility that exploratory decisions might utilize previously learned rules and associations. Many tasks which investigate the explore/exploit dilemma are well learned and induce exploration by altering reward delay (Hayden et al., 2011), magnitude (Daw et al., 2006), or probability (Behrens et al., 2007; Knox et al., 2012). What is unclear from these tasks is the degree to which animals use learned rules and environmental models to guide their exploration, and how animals might explore in a novel circumstance.

If animals do not generalize learned rules to novel circumstances, what does control exploratory actions, and how do these actions relate to exploitation? The explore/exploit dilemma is classically characterized as a direct trade-off (Cohen et al., 2007). You are either exploring or exploiting, and doing one necessarily precludes the other. Tasks like the n-armed bandit have reinforced this view, where the mathematically optimal decision (to maximize reward) is defined as “exploit” while all other choices are “explore” (Daw et al., 2006). But such a forced choice is rare in the real world. While actions controlled by exploration and exploitation decision processes cannot occur simultaneously, outside of the lab there are often many choice options available that do not explicitly fall into “exploration” or “exploitation”. This raises the possibility that the decision-making aspects of exploration and exploitation run in parallel and do not necessarily directly compete. Thus, it is unclear both the extent to

which exploration utilizes information gleaned from the environment, and if and how exploration and exploitation directly compete.

While a large body of work focuses on the explore/exploit dilemma in relation to contextual and cued information, action control may rely on similar processes. The prefrontal cortex has substantial evidence implicating it in learning and applying rules (Badre et al., 2010; Wallis et al., 2001; White & Wise, 1999) in mediating the explore/exploit dilemma (Beharelle et al., 2015; Boorman et al., 2009; Cohen et al., 2007; Daw et al., 2006; Laureiro-Martínez et al., 2014; Morris et al., 2016) and in action control (Balleine & O'Doherty, 2009). For example, the anterior cingulate cortex has been strongly implicated in the explore/exploit dilemma (Hayden et al., 2011), while orbital frontal cortex (OFC) and secondary motor cortex (M2) have been implicated in controlling goal-directed instrumental actions (Gremel & Costa, 2013a, 2013b). It may be that cortical circuits underlying action control could be differentially recruited during explore and exploit processes. Within this framework, OFC has been shown to be necessary for actions sensitive to changing action value (Gourley et al., 2016; Gremel et al., 2016; Gremel & Costa, 2013b; Rhodes & Murray, 2013) and partially observable states (Bradfield et al., 2015). M2 has been shown to support goal-directed actions (Gremel & Costa, 2013a) and the contingency between actions (Ostlund et al., 2009; Siniscalchi et al., 2016; Yin, 2009). OFC and M2 regions are reciprocally connected (Zingg et al., 2014), but not onto overlapping populations (i.e. OFC terminal fields in M2 do not overlap with M2 somata that project to OFC, and vice versa) (Oh et al., 2014). Furthermore, structural plasticity of OFC projections into M2 (OFC-M2) correlates with rule learning (Johnson et al., 2016) – specifically, bouton

gain correlates with rule learning and subsequent exploitation, while bouton loss correlates with exploration. This suggests that OFC-M2 projections could contribute to or occlude exploration following rule learning.

We used a self-paced operant instrumental lever press task in mice to determine if exploration utilizes learned rules and the extent to which exploration and exploitation directly compete. In this task (Hilario et al., 2007, 2012; Iguchi et al., 2017), mice are trained to press one lever for a food reward. Then during the test session a novel but perceptually similar lever is also inserted into the chamber, and we measure responses on the trained and novel levers. Different schedules of reinforcement can be used to bias either exploitation of the trained lever or exploration of the novel lever (Hilario et al., 2007, 2012). Previous studies using this particular task have hypothesized that responding reflects either exploration (Hilario et al., 2007; Iguchi et al., 2017) or action generalization mechanisms (Hilario et al., 2012), though this has not been tested.

We first probed the ability for outcome value to affect responding on the novel lever, and found no evidence that changes in outcome value affect novel lever exploration. Next, we evaluated if temporal uncertainty would affect exploration, and again found no evidence to suggest that temporal uncertainty affects novel lever exploration. Correlative data revealed that the amount of experience mice had with the learned action-outcome rule correlated with exploitation of the trained lever. Importantly, experience did not correlate – either positively or negatively – with exploration. That is, roughly the same level of exploration occurred irrespective of how much experience mice had with the learned rule, indicating that the decision-making

processes that mediate exploration and exploitation may not *directly* compete (i.e., more exploitation does not *necessarily* mean less exploration in a free operant context). This led us to examine OFC-M2 projection neurons which, as mentioned, are involved in rule learning (Johnson et al., 2016). Inhibition of OFC-M2 projection neurons during training and testing increased exploration and reduced exploitation. Overall our data suggest that mice do not generalize previously learned rules when engaging in novel lever exploration, that exploitation and exploration decision processes may run in parallel, and that the OFC-M2 circuit is a critical node controlling the emergence of exploitative action control.

Results

Outcome devaluation does not affect lever generalization

We first examined whether mice generalize sensory-specific food outcome expectancies to the novel lever. We took advantage of two different schedules of reinforcement, with a random ratio (RR) schedule biasing sensitivity to sensory-specific changes in food value and a random interval (RI) schedule biasing relative insensitivity to value changes (Dickinson, 1985; Dickinson et al., 1983). Previous work has found that RR schedules also bias more exploitation of the trained lever while RI schedules bias increased exploration of the novel lever (Hilario et al., 2007, 2012). Hence, if mice are generalizing sensory-specific features of the expected food outcome, then outcome devaluation should produce decreased exploratory pressing of the novel lever under an RR schedule in comparison to a RI schedule.

Mice were trained to press a lever located left or right of a food magazine (counterbalanced) for food pellets under either a RR or RI schedule. Response requirement increased across training, with RI schedules progressing from RI 30s to RI 60s, and RR10 progressing to RR20 after two days of schedule training (Figure 1.1a). Mice trained under a RR schedule increased their response rate across training to a greater degree than those trained under a RI schedule (Figure 1.1b). A two-way repeated-measures ANOVA (Day x Schedule) performed on acquisition response rate (lever presses/minute) revealed a significant interaction ($F_{(16,224)} = 5.22$, $p < 0.0001$) and significant main effects of Day ($F_{(16,224)} = 17.5$, $p < 0.0001$) and Schedule ($F_{(1,14)} = 19.9$, $p = 0.0005$), with post-hoc analyses (Bonferroni corrected) showing schedules differed on most of the training days.

We then performed an outcome devaluation procedure counterbalanced across two days, where the operant outcome is devalued using sensory-specific satiety on the devalued (DV) day, while on the valued (V) day an outcome previously experienced in the homecage is pre-fed to control for effects of general satiation. Following 1 hour free feed access to either the operant or homecage outcome, mice were placed in the operant chamber for a 5 minute extinction test. On both the V and DV day, a second novel lever was inserted (either left or right of the food magazine, counterbalanced) in addition to the trained lever. Mice were re-trained for one day in between the V and DV day.

Outcome devaluation procedures had no effect on exploration of the novel lever in mice trained either on RR or RI schedules (Figure 1.1c). A three-way repeated-measures ANOVA (Lever Type x Valuation State x Schedule) showed a significant

three-way interaction ($F_{(1,9)}$ = 14.6, p = 0.004). A significant two-way interaction between Schedule and Lever Type ($F_{(1,9)}$ = 11.7, p = 0.008), showed schedule–induced differences in exploration/exploitation as previously observed (Hilario et al., 2007, 2012). There was also a significant interaction between Lever Type and Valuation State ($F_{(1,9)}$ = 19.4, p = 0.002), indicating that, overall, only the Trained lever was sensitive to value manipulations. There was no interaction between Schedule and Valuation State ($F_{(1,9)}$ = 3.22, p = 0.11). Main effects of Schedule ($F_{(1,9)}$ = 19.7, p = 0.002), Lever type ($F_{(1,9)}$ = 27.7, p < 0.001), and Valuation State ($F_{(1,9)}$ = 8.29, p = 0.02) were also observed. Planned post-hoc comparisons (Bonferroni corrected) between V and DV days were made for each Lever by Schedule combination. Devaluation significantly reduced Trained lever pressing in RR-trained mice ($t_{(8)}$ = 3.33, p = 0.01), but had no effect on Trained lever pressing in RI-trained mice (p = 0.23). Devaluation had no effect on Novel lever pressing in either RR (p = 0.71) or RI (p = 0.52) trained mice.

To determine if a conditioned context-outcome association influenced performance, we also measured head-entries into the magazine. We found no effect of outcome devaluation on the conditioned head-entry response (Figure 1.1d). A two-way RM ANOVA (Valuation State x Schedule) showed no significant interaction between Valuation State and Schedule ($F_{(1,9)}$ = 0.303, p = 0.60), nor a significant main effect of Valuation State ($F_{(1,9)}$ = 2.76, p = 0.13), although there was a main effect of Schedule ($F_{(1,9)}$ = 16.3 p = 0.003). Thus, outcome devaluation does not seem to reduce head-entries, suggesting that the context-outcome pairing was not significantly devalued following satiation procedures.

In addition, differences in conditioned response rates acquired between schedules did not contribute to these results (Figure 1.S1). We performed linear regression analyses on average response rate by Devaluation Index (DV index, see methods) to compare the relationship between response rates during training to the degree of outcome devaluation. There was no significant relationship when comparing late acquisition response rate and DV index on the trained ($F_{(1,9)} = 2.96$, $p = 0.12$; $R^2 = 0.25$) or novel ($F_{(1,9)} = 0.52$, $p = 0.49$; $R^2 = 0.055$) lever. Similarly, there was no significant relationship between early response rate and DV index on either the trained or novel lever (Figures 1.S1 a-b). Since novel lever presses were lower than trained lever presses, there is the possibility that floor effects could prevent mice from decreasing their novel presses following devaluation. We ran linear regressions of lever press rate during testing on the trained and novel levers with DV Index (for the respective lever). We found no correlation between press rate on the Valued day and DV index for either the trained or novel lever (Figure 1.S1c). Likewise, we found no correlation between the average press rate across Valued and Devalued days and DV index for either the trained or novel levers (Figure 1.S1d). Hence we found no evidence that response rate during either acquisition or test contributes to the magnitude of outcome devaluation. Outcome devaluation does not appear to affect novel lever exploration, and this was true in mice trained in either a RR or RI schedule, which bias sensitivity or insensitivity (respectively) of trained lever pressing to outcome devaluation.

Figure 1.1

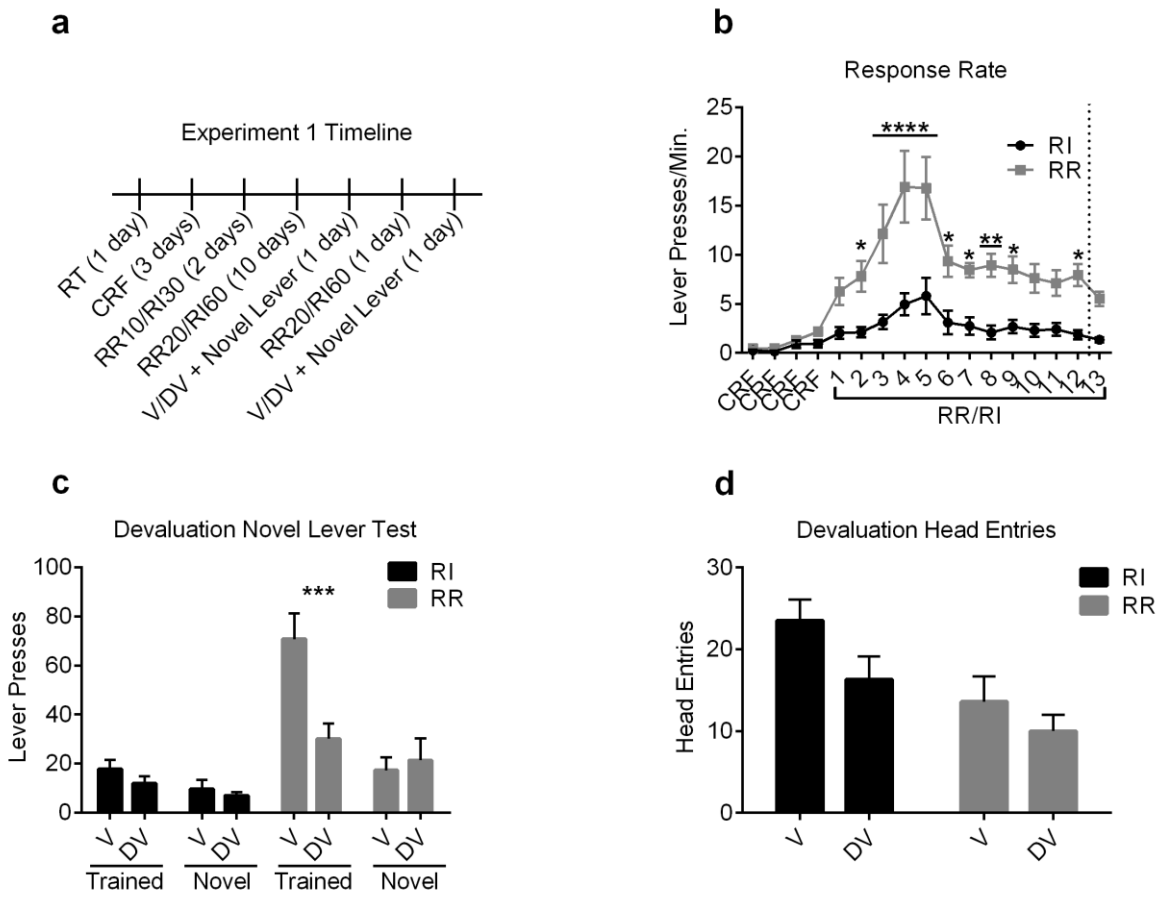
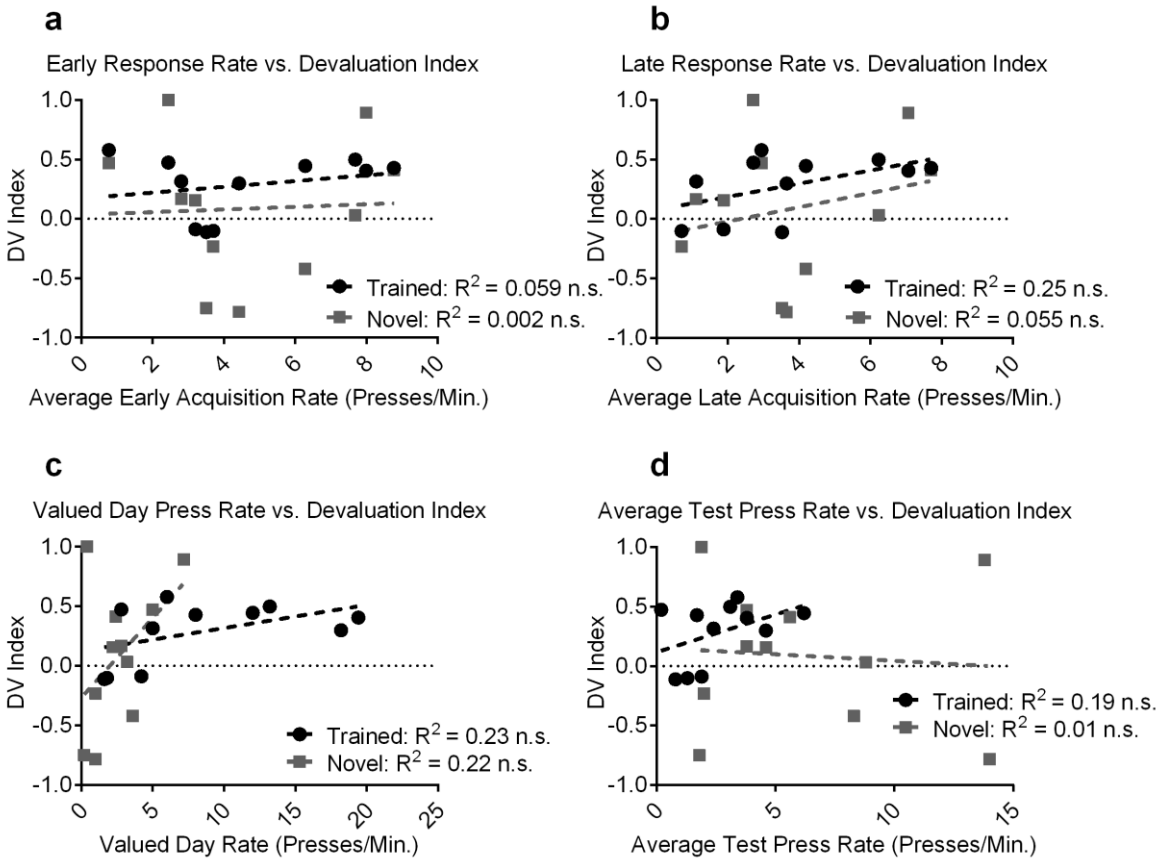


Figure 1.1. Outcome value does not contribute to novel lever pressing. Mice were traine=d to press a lever for an outcome under a random ratio (RR) or random interval (RI) schedule and then underwent a combined outcome devaluation/novel lever test. **(a)** Experimental timeline. **(b)** Response rate (Lever Presses/Min.) during acquisition. Days 1-2 were conducted under a RR10/RI30 schedule, remaining days were under a RR20/RI60 schedule. Dotted line indicates where first test day occurred, followed by one day of re-training and then the second test day. Significance markers indicate post-hoc differences between schedules. **(c)** Combined devaluation novel lever test. **(d)** Head entries into the magazine during the combined devaluation novel lever test. RT = Random Time. CRF = Continuous Ratio of Reinforcement. V = Valued Day. DV = Devalued Day. V/DV + Novel Lever = Combined Devaluation Novel Lever Test. Error Bars = ± SEM. n.s. = Not Significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

Supplemental Figure 1.S1



Supplemental Figure 1.S1. No correlation between acquisition or test response rate and sensitivity to devaluation. We ran several linear regressions comparing either acquisition (**a-b**) or test (**c-d**) response rate to Devaluation Index (DV Index) to determine if response rates influenced sensitivity to devaluation for either lever. **(a)** Early acquisition response rate (average of the first two days of schedule training) vs. DV Index. Neither the trained ($F_{(1,9)} = 0.57$, $p = 0.47$; $R^2 = 0.059$) nor the novel ($F_{(1,9)} = 0.02$, $p = 0.89$; $R^2 = 0.002$) lever slope differed from 0. **(b)** Late acquisition response rate (final three day average) vs. DV Index. Neither the trained ($F_{(1,9)} = 2.96$, $p = 0.12$; $R^2 = 0.25$) nor the novel ($F_{(1,9)} = 0.52$, $p = 0.49$; $R^2 = 0.055$) lever slope differed from 0. **(c)** Response rate during the Valued test day vs. DV Index for the respective lever. We found no correlation between press rate on the Valued day and DV index for either the trained ($F_{(1,9)} = 2.68$, $p = 0.14$; $R^2 = 0.229$) or novel lever ($F_{(1,9)} = 2.57$, $p = 0.14$; $R^2 = 0.222$). **(d)** Average response rate during test (Valued and Devalued days) vs. DV Index for the respective lever. We found no correlation between the average press rate across Valued and Devalued days and DV index for either the trained ($F_{(1,9)} = 2.06$, $p = 0.19$; $R^2 = 0.186$) or novel lever ($F_{(1,9)} = 0.06$, $p = 0.81$; $R^2 = 0.007$). Dotted linear regression lines indicate non-significant slopes (compared to 0). n.s. = Not Significant.

Uncertainty does not affect action generalization

Uncertainty is known to modulate the balance between exploration and exploitation (Cohen et al., 2007). Since previous work has shown that increasing temporal uncertainty (i.e., uncertainty regarding *when* a reward is available) in RI schedules biases the development of habitual actions (Derusso et al., 2010), and RI schedules promote generalization (Hilario et al., 2007, 2012), we hypothesized that increases in temporal uncertainty might lead to increased exploration of the novel lever.

Mice were trained under three different schedules (Figure 1.2a) that differed in terms of their reward probability distribution, but shared the same average time to reward (Figure 1.2b). This was achieved by utilizing different time cycles (T) coupled with different probabilities (p). In the Fixed Interval 60s schedule (FI60), T = 60s and p = 1.0, such that at every 60s cycle, there is 100% chance of a reinforcer being earned following a lever press. In the Random Interval 60s (p = 0.5) schedule, T = 30s and p = 0.5, such that at every 30s cycle, there is a 50% chance of a press producing a reinforcer. In the Random Interval 60s (p = 0.1) schedule, T = 6s and p = 0.1, such that at every 6s cycle, there is a 10% chance of a press producing a reinforcer. Importantly, the average time to reward is 60s in all three schedules (Figure 1.2b). These schedules did not produce different response rates during acquisition (Figure 1.2c), as evidenced by a two-way repeated measures ANOVA (Day X Schedule) that showed no interaction ($F_{(20,420)}$ = 0.64, p = 0.89) or main effect of Schedule ($F_{(2,42)}$ = 0.25, p = 0.78), but did show a main effect of Day ($F_{(10,420)}$= 38.7, p < 0.0001). We confirmed that our manipulation led to changes in action-outcome contiguity (the average time

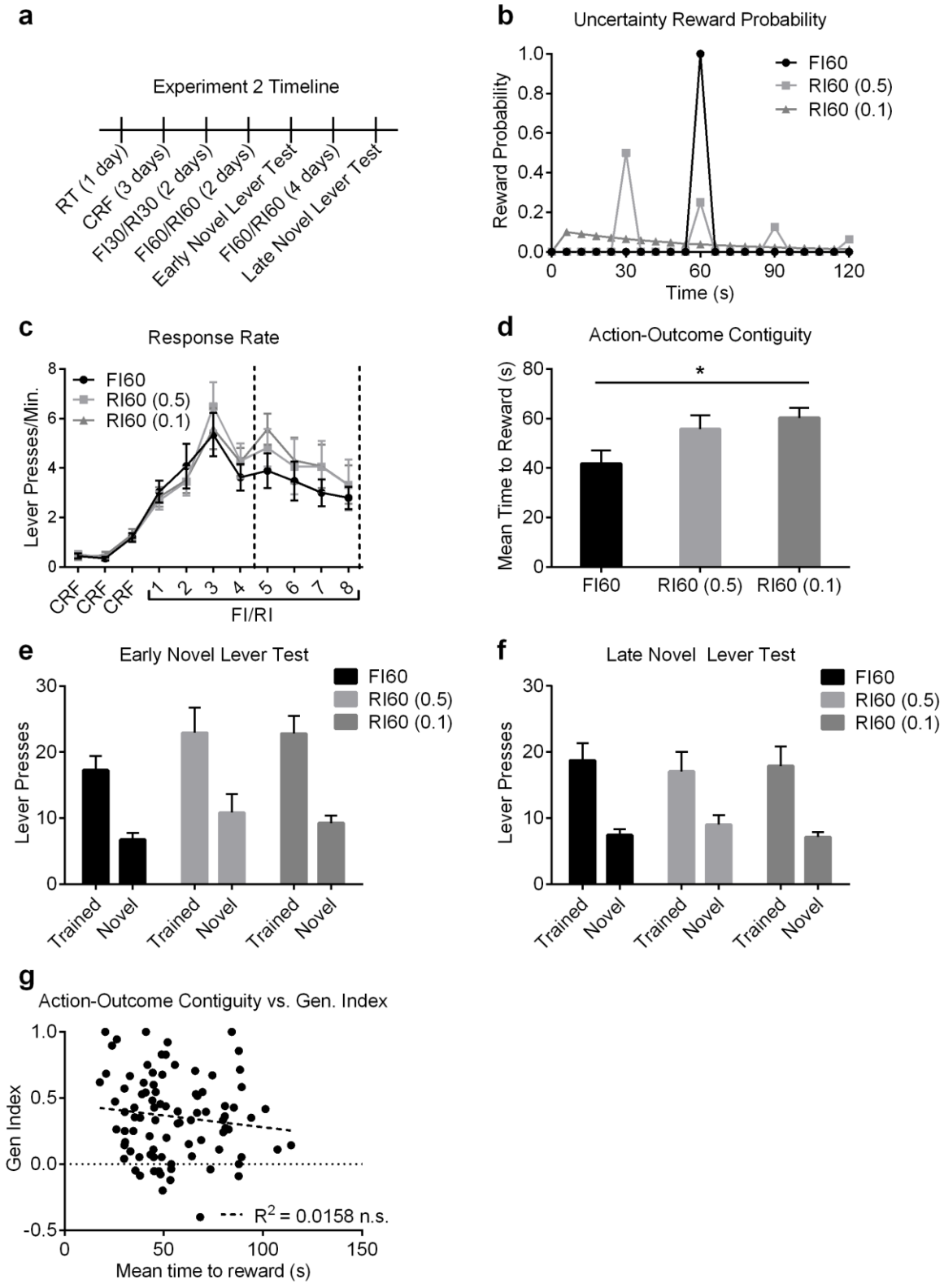
between a lever press and an outcome delivery) (Derusso et al., 2010) on the last acquisition day prior to the first novel lever test (one-way ANOVA; significant effect of schedule ($F_{(2,41)} = 3.86$, $p = 0.029$) (Figure 1.2d). Hence mice learned to press the lever under different degrees of temporal uncertainty.

We found no evidence to suggest that temporal uncertainty affects exploration of the novel lever. Mice were given two novel lever tests where an additional, novel lever was inserted into the chamber along with the trained lever; an early test was conducted after initial acquisition at a time point early on in rule learning, and a second late test was conducted after extended training, although in this case the additional lever was not completely novel. A two-way repeated measures ANOVA (Lever Type X Schedule) conducted on lever presses in the early test did not show an interaction ($p = 0.77$) or a main effect of Schedule ($p = 0.16$), but did show a main effect of Lever Type ($F_{(1,42)} = 47.7$, $p < 0.0001$) (Figure 1.2e). Similarly, a two-way repeated measures ANOVA conducted on lever pressing during the late test did not show an interaction ($p = 0.73$) or main effect of Schedule ($p = 0.96$), but did show a significant main effect of Lever Type ($F_{(1,42)} = 33.5$, $p < 0.0001$) (Figure 1.2f). As these three interval schedules have been demonstrated to differ in their action-outcome contiguity (Derusso et al., 2010) (Figure 1.2d), we correlated action-outcome contiguity with Generalization Index (Gen. Index: values close to 1 indicate complete exploitation of the trained lever, while values near 0 indicate generalized responding to both levers, see methods). We found no correlation between the action-outcome contiguity on the last training day and the degree to which mice generalized lever pressing to the novel lever during testing ($F_{(1,88)} = 1.40$, $p = 0.24$; $R^2 = 0.02$) (Figure 1.2g). Overall, our data show mice exhibited

weak generalization of responding, and we found no evidence that temporal uncertainty influenced novel lever exploration.

Figure 1.2. Uncertainty does not contribute to novel lever pressing. Mice were trained to press a lever for an outcome under one of three different interval schedules which varied in their uncertainty. **(a)** Experimental timeline. **(b)** Reward distribution of the three different interval schedules. Note that while the temporal distribution of reward availability differs, all three schedules share the same average time to reward (60s). **(c)** Response rate during acquisition. Dotted lines indicate where novel lever tests occurred. **(d)** Action-outcome contiguity, defined as mean time between a lever press and reward on the final acquisition day prior to the first novel lever test. **(e)** Early and (**f)** late novel lever test lever presses. In both graphs there is a significant main effect of lever. **(g)** Correlation between action-outcome contiguity and generalization index (Gen. Index), calculated as (Trained Presses – Novel Presses) / Total Presses. FI60 is a Fixed Interval 60s schedule. RI60 $p = 0.5$ is a Random Interval 60s schedule with moderate uncertainty. RI60 $p = 0.1$ is a Random Interval 60s schedule with high uncertainty. RT = Random Time. CRF = Continuous Ratio of Reinforcement. Error Bars = ± SEM. n.s. = Not Significant, * = $p < 0.05$.

Figure 1.2


a
Experiment 2 Timeline
RT (1 day)
CRF (3 days)
FI30/RI30 (2 days)
FI60/RI60 (2 days)
Early Novel Lever Test
FI60/RI60 (4 days)
Late Novel Lever Test
b
Uncertainty Reward Probability
Reward Probability
Time (s)
FI60
RI60 (0.5)
RI60 (0.1)
c
Response Rate
Lever Presses/Min.
CRF
FI/RI
d
Action-Outcome Contiguity
Mean Time to Reward (s)
e
Early Novel Lever Test
Lever Presses
Trained
Novel
f
Late Novel Lever Test
g
Action-Outcome Contiguity vs. Gen. Index
Gen Index
R^2 = 0.0158 n.s.
Mean time to reward (s)

Action Experience Biases Selective Exploitation

We next sought to determine if the amount of experience with the learned action biased towards exploitation, as previously reported (Iguchi et al., 2017). Utilizing data obtained from the mice in the uncertainty experiment above, we calculated the total lever presses made since the start of schedule training until either the early or late generalization test. We found that experience with the learned action did indeed bias towards exploitation. A linear regression analysis of total lever presses during acquisition and the generalization index revealed a small but significant positive relationship ($F_{(1,88)} = 8.43$, $p = 0.005$; $R^2 = 0\ .087$) (Figure 1.3a), with more total lever presses during acquisition leading to higher generalization index values (i.e., more exploitation). We ran separate linear regressions broken up by training schedule (FI vs. RI (0.5) vs. RI (0.1)) to determine if this effect was primarily driven by one schedule. We found that there was still a significant relationship between total lever presses during acquisition and generalization index in the FI ($F_{(1,28)}= 6.67$, $p = 0.015$; $R^2 = 0.19$), and the RI(0.1) ($F_{(1,32)}= 5.88$, $p = 0.02$; $R^2 = 0.16$) schedules, but not in the RI(0.5) schedule ($F_{(1,26)}= 0.03$, $p = 0.86$; $R^2 = 0.0013$) (Figure 1.S2a). This demonstrates that this relationship is not driven by only one schedule, and indeed is observed in the schedules that differ most in terms of their uncertainty (that is, uncertainty does not appear to contribute to the correlation between experience and exploitation).

An increased generalization index could indicate either an increase in trained lever presses and/or a decrease in novel lever presses. We therefore ran linear regressions using total lever presses during acquisition by trained or novel lever

presses collapsed across early and late tests (Figure 1.3b). Interestingly, we found a significant relationship with only trained lever presses ($F_{(1,88)} = 18.5$, $p < 0.0001$; $R^2 = 0.17$), and not with novel lever presses ($F_{(1,88)} = 0.07$, $p = 0.79$; $R^2 = 7.97e\text{-}4$). Furthermore, the slope of these two lines (trained vs. novel lever press) differed significantly ($F_{(1,176)} = 15.1$, $p = 0.0001$), indicating that the amount of experience with the trained lever is highly predictive of trained lever presses on test, but does not impact the degree of novel lever exploration. Indeed, this relationship was present on the last day of training prior to testing, where we again find a significant relationship between last day response rate and generalization index ($F_{(1,88)}= 17.2$, $p < 0.0001$; $R^2 = 0.16$) (Figure 1.3c), and with test response rates on the trained ($F_{(1,88)}= 133$, $p < 0.0001$; $R^2 = 0\ .60$) but not the novel ($F_{(1,88)}= 3.54$, $p = 0.06$; $R^2 = 0.04$) lever, and again the slopes of these two lines differed significantly ($F_{(1,176)} = 62.5$, $p < 0.0001$) (Figure 1.3d).

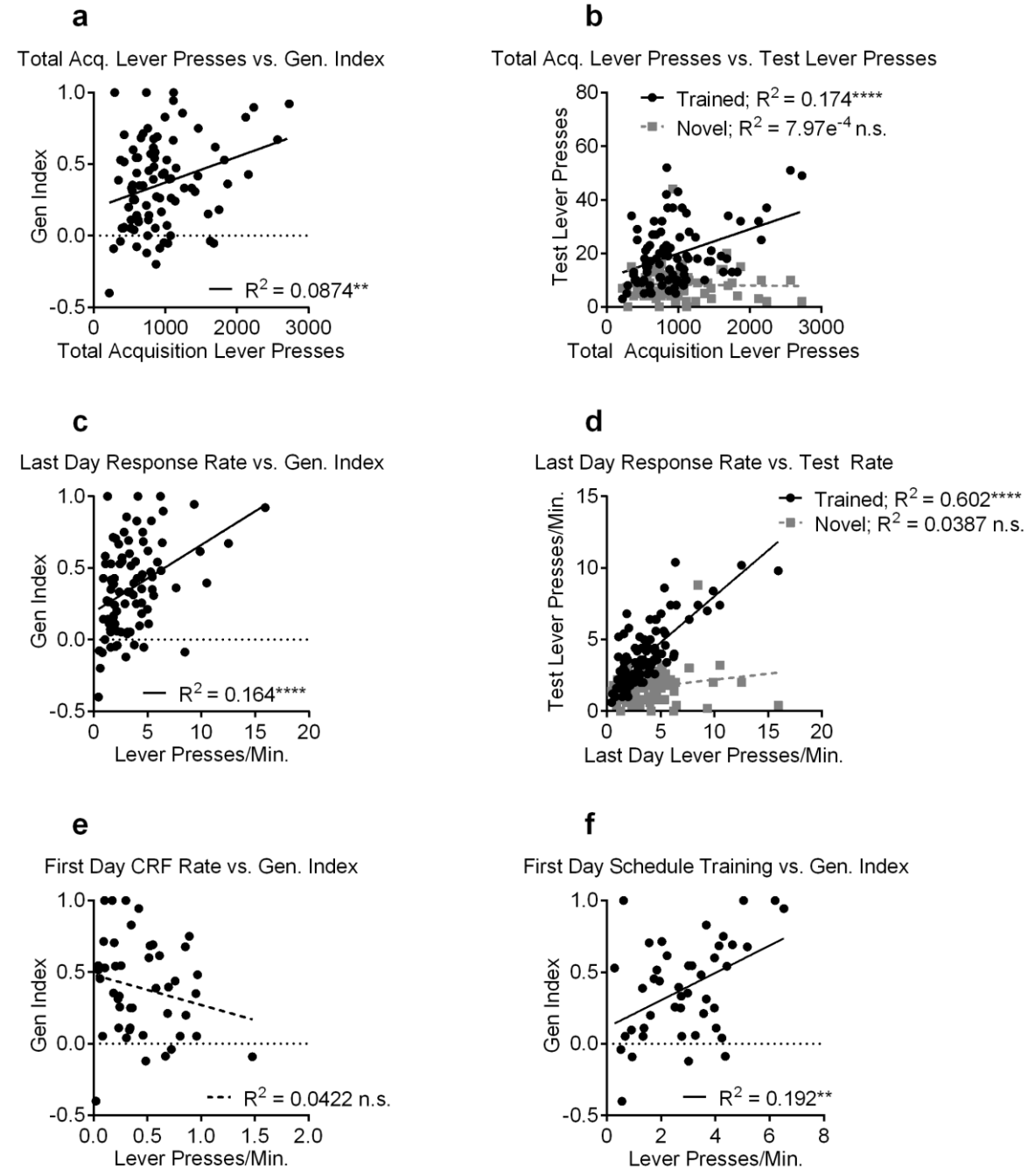
We next sought to determine how early this relationship between response rate and generalization index emerged. For these analyses, we used data only from the early generalization test to examine the relationship between initial learning and testing. Using response rates from the very first day of CRF (Continuous Ratio of Reinforcement) training, we found no significant relationship with the subsequent generalization index ($p = 0.18$, $R^2 = 0.04$) (Figure 1.3e). This lack of a significant relationship persisted throughout the following 2 days of CRF training (Figures 1.S2a-c), though it should be noted that the low response rates during this initial CRF training might make correlations difficult to detect. However, by the first day of schedule training on FI30 or RI30, a significant relationship between response rate and the

generalization index emerged ($F_{(1,43)} = 10.2$, $p = 0.003$; $R^2 = 0.19$) (Figure 1.3f). This suggests that differences in the action-outcome relationships experienced during early schedule learning contribute to exploitation on the trained lever.

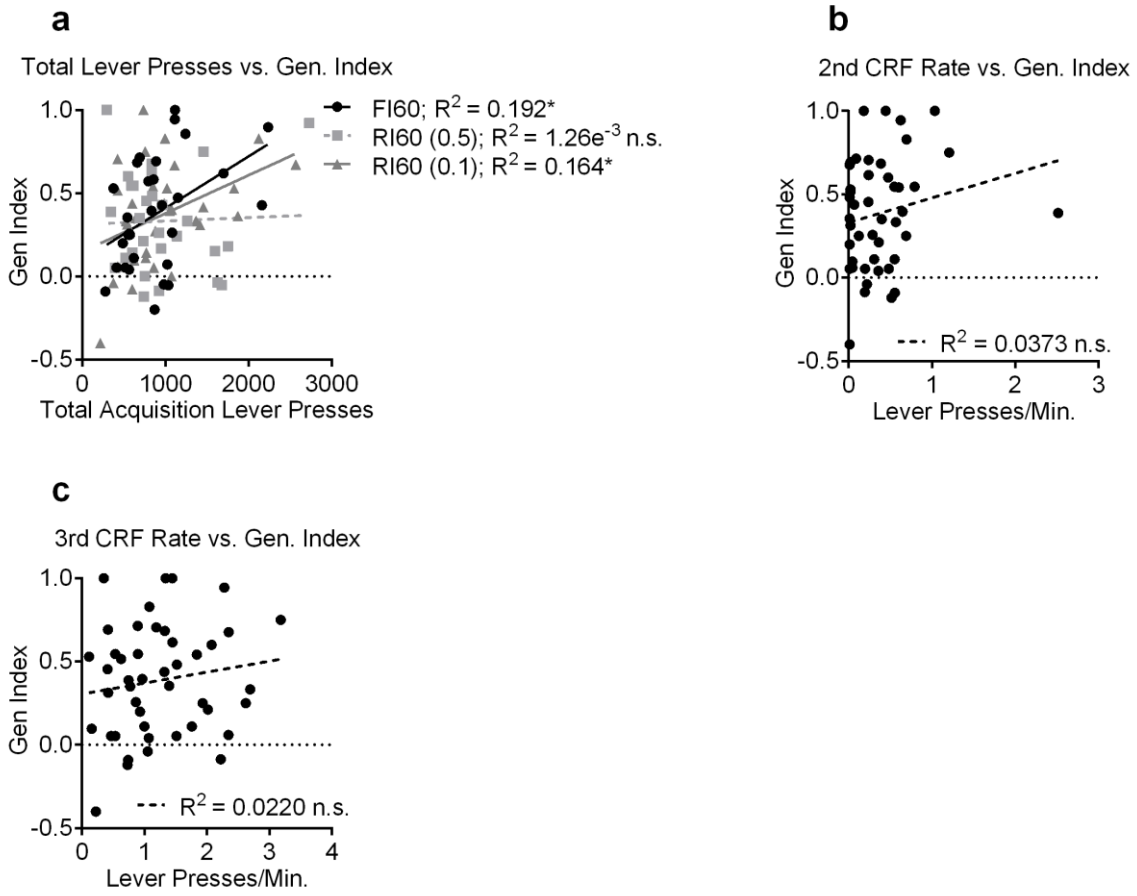
These results indicate that the amount of experience with a known action-outcome relationship is predictive of the subsequent degree of exploitation during a probe test, with more experience and higher rates of responding correlating with increased exploitation of the trained lever. However, there was no correlation with exploration, as might be expected if actions were being generalized. Similarly, if exploitation and exploration decision-processes directly competed with one another, we should expect to see a negative correlation (that is, as exploitation increases with experiences, exploration should decrease), but instead we see no relationship between experience and exploration whatsoever. When we measured the duration mice hold the trained versus novel lever down in a separate cohort of mice, we found that lever press durations can differ between trained and novel levers (Figures 1.S3a-d), indicating that the motor response itself may not fully generalize. Together, the results of our uncertainty experiment provide evidence that the learned Stimulus-Response association does not generalize to the novel lever.

Figure 1.3. Experience with the trained lever correlates with exploitation but not exploration. The same mice from Figure 1.2 were used to run these correlations. **(a)** Correlation between total lever presses during acquisition and generalization index (Gen. Index). **(b)** Correlation between total lever presses during acquisition and test lever presses on the trained or novel lever. **(c)** Correlation between last day response rate and generalization index. **(d)** Correlation between last day response rate and test response rate on trained and novel levers. **(e)** Correlation between response rate on the final CRF (Continuous Ratio of Reinforcement) training day and generalization index. **(f)** Correlation between response rate on the first day of schedule training and generalization index. Dotted linear regression lines indicate non-significant correlations, while solid linear regression lines are significant. Acq. = Acquisition. n.s. = Not Significant, ** = $p < 0.01$, **** = $p < 0.0001$.

Figure 1.3

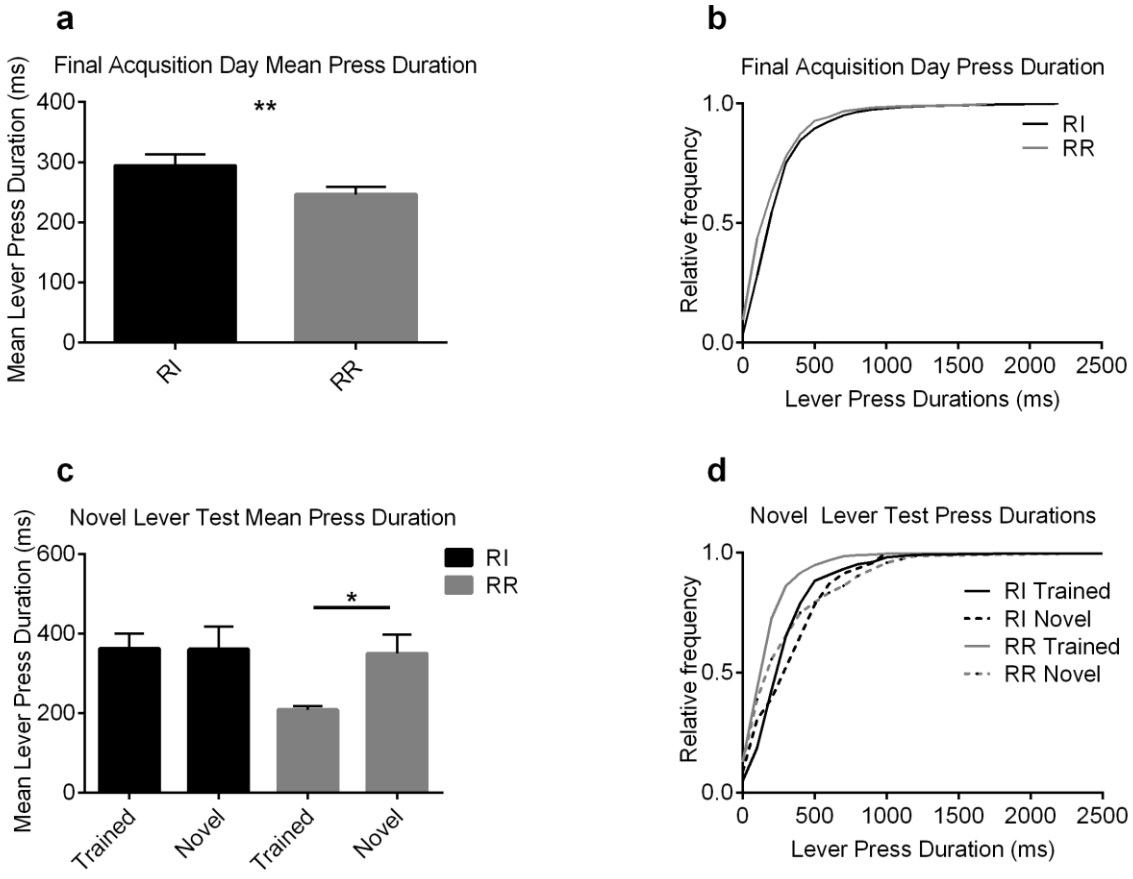

a
Total Acq. Lever Presses vs. Gen. Index
Gen Index
1.0
0.5
0.0
-0.5
0
1000
2000
3000
Total Acquisition Lever Presses
R2 = 0.0874**
b
Total Acq. Lever Presses vs. Test Lever Presses
Test Lever Presses
80
60
40
20
0
Trained; R2 = 0.174****
Novel; R2 = 7.97e-4 n.s.
Total Acquisition Lever Presses
c
Last Day Response Rate vs. Gen. Index
Gen Index
Lever Presses/Min.
R2 = 0.164****
d
Last Day Response Rate vs. Test Rate
Test Lever Presses/Min.
15
10
5
0
Trained; R2 = 0.602****
Novel; R2 = 0.0387 n.s.
Last Day Lever Presses/Min.
e
First Day CRF Rate vs. Gen. Index
R2 = 0.0422 n.s.
f
First Day Schedule Training vs. Gen. Index
R2 = 0.192**

Supplemental Figure 1.S2



Supplemental Figure 1.S2. Correlation between experience and exploitation exists across different schedules and is not present prior to schedule training. **(a)** Correlation between total lever presses during acquisition and generalization index (Gen. Index) separated by schedule. FI60 = Fixed Interval 60s. RI60 (0.5) is a Random Interval 60s schedule with moderate uncertainty. RI60 (0.1) is a Random Interval 60s schedule with high uncertainty. **(b)** Correlation between rate on the second day of CRF (Continuous Ratio of Reinforcement) training and generalization index. **(c)** Correlation between rate on the third day of CRF training and generalization index. Dotted linear regression lines indicate non-significant slopes (compared to 0), while solid lines indicate significant slopes. n.s. = Not Significant, * = $p < 0.05$.

Supplemental Figure 1.S3



Supplemental Figure 1.S3. Lever press duration differs between trained and novel levers in ratio-trained mice. To examine how similar action performance itself (the response) is on the trained versus novel lever, mice were trained on either a Random Ratio (RR) or Random Interval (RI) schedule and the duration of their lever presses was recorded. **(a)** Mean lever press duration on the final day of acquisition. A Mann-Whitney test of lever durations on the final day of training prior to testing revealed significant differences between the two schedules (Mann-Whitney U = 29523, n_{RI}= 202; n_{RR}= 305, p = 0.002). **(b)** Distribution of lever press durations on the final day of acquisition. **(c)** Mean lever press durations on the trained and novel lever during the novel lever test. While mice trained under a RI schedule made trained and novel lever presses of similar durations (Mann-Whitney U = 29523, $n_{Trained}$ = 204; n_{Novel} = 23, p = 0.64), those trained under a RR schedule pressed the trained and novel levers with different durations (Mann-Whitney U = 13782, $n_{Trained}$ = 462; n_{Novel} = 72, p = 0.02) **(d)** Distribution of lever press durations during the novel lever test. Bars = ± SEM. n.s. = not significant, * = $p < 0.05$, ** = $p < 0.01$.

Orbital frontal cortex projections to secondary motor cortex mediate learned action-outcome associations

Our data suggests that increased experience drives exploitation of known rules. Rule learning in uncertain environments has been proposed to induce structural plasticity of OFC terminals in M2, with the magnitude of this plasticity correlating with subsequent exploitation of known rules (Johnson et al., 2016). We hypothesized that activity of OFC projections to M2 is necessary for rule learning that supports exploitation of the trained lever. Hence, inhibiting OFC projections to M2 during both learning and testing should occlude this plasticity, and thereby bias exploration during the novel lever test.

We utilized a dual viral vector approach to isolate OFC projections into M2, and used chemogenetics to specifically attenuate OFC-M2 activity (Figure 1.4a). Mice were given bilateral injections in OFC of a rAAV5/hSyn-DIO-hM4D-mcherry expressing a Cre-dependent inhibitory Designer Receptor Exclusively Activated by a Designer Drug (DREADD) (Armbruster et al., 2007) or a rAAV5/hSyn-DIO-mcherry expressing a Cre-dependent fluorophore control (mCherry). In M2, all mice received bilateral injections of AAV5/CamKIIα-GFP-Cre expressing GFP-Cre under the control of the CamKIIα promoter that can be transferred retrograde (Rothermel et al., 2013). We observed minimal expression of neurons which project in the other direction (M2 to OFC: as evidenced by lack of mCherry in M2 and lack of GFP in OFC; Figure 1.4b).

All mice were trained under a RI schedule. All animals received injections of the hM4D agonist CNO (1.0mg/ml) 30 minutes prior to all schedule training and test days,

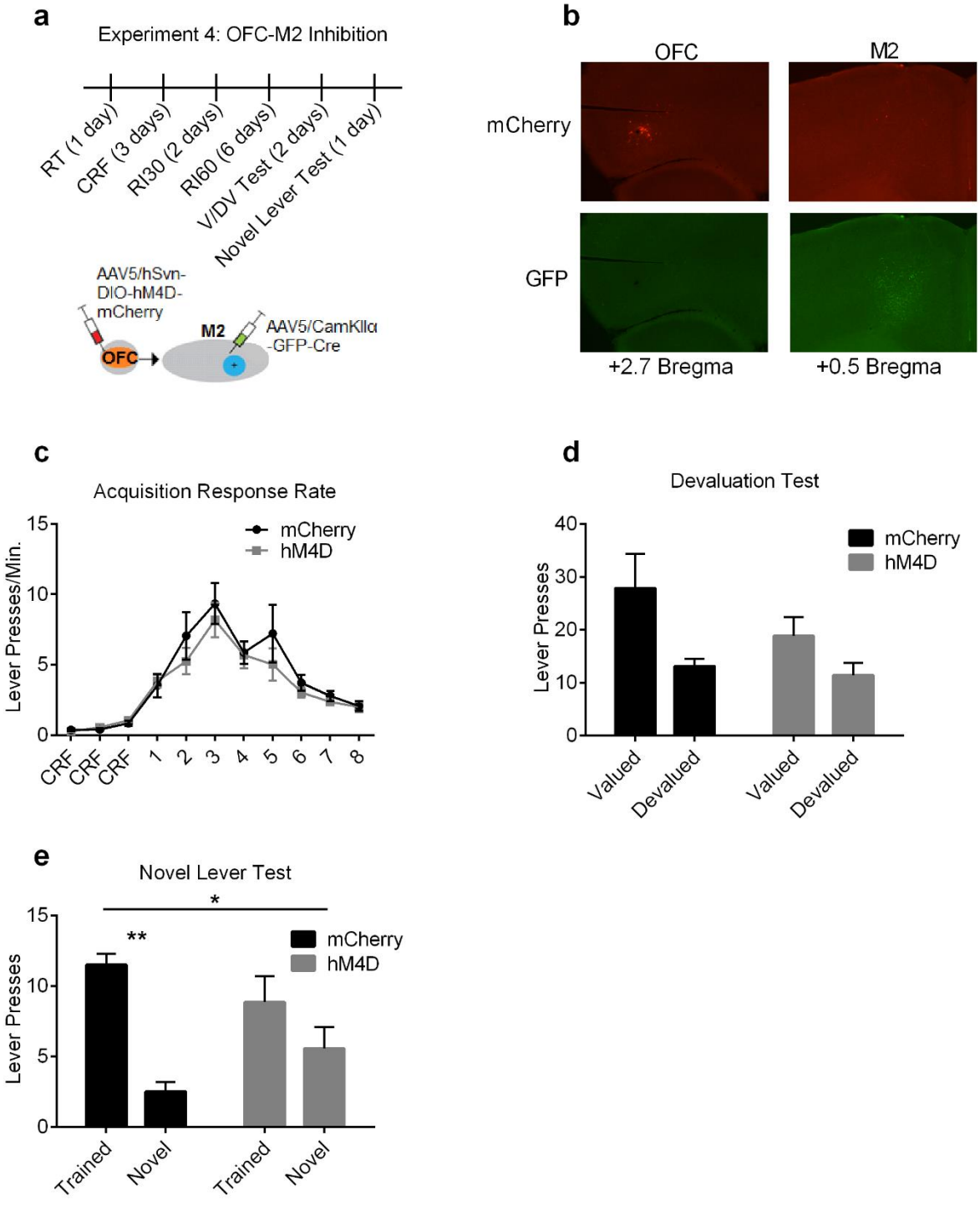
a duration we have previously shown sufficient to reduce OFC cell excitability (Gremel et al., 2016; Gremel & Costa, 2013b). hM4D and mCherry control mice showed similar acquisition of lever press behavior (Figure 1.4c). A two-way repeated measures ANOVA (Day x Virus) did not show a significant interaction ($p = 0.82$), or main effect of Virus ($p = 0.46$), but did show a main effect of Day ($F_{(10,130)} = 25.3$, $p < 0.0001$). Since both OFC (Gremel & Costa, 2013b) and M2 (Gremel & Costa, 2013a) are individually necessary for goal-directed actions under outcome devaluation, we first sought to test if the projections from OFC to M2 were specifically necessary for goal-directed actions. We took advantage of previous findings that action control relatively early in training under RI schedules is still goal-directed (Dickinson et al., 1995; Shan et al., 2014), and performed outcome devaluation procedures after relatively little training. A two-way repeated measures ANOVA (Valuation state x Virus treatment) showed no interaction ($p = 0.31$), nor a main effect of virus ($p = 0.26$). Only a main effect of Valuation State ($F_{(1,13)} = 10.1$, $p = 0.007$) was observed, indicating that OFC-M2 activity attenuation during training and testing did not disrupt goal-directed control (Figure 1.4d).

Following devaluation testing, we next assessed the involvement of the OFC-M2 projection in a second test session in which the novel lever was introduced. In contrast to our outcome devaluation results, we found that attenuation of OFC-M2 projection neuron activity decreased exploitation of the trained lever in relation to exploration of the novel lever (Figure 1.4e). While mCherry control mice pressed the trained lever to a much greater degree than the novel lever, hM4D mice pressed each of the levers a similar amount of times. A two-way repeated measures ANOVA (Lever x Virus), revealed a significant interaction ($F_{(1,13)} = 5.97$, $p = 0.03$) and significant main

effect of Lever ($F_{(1,13)} = 27.6$, $p = 0.0002$), but no main effect of Virus ($p = 0.87$). Bonferroni-corrected post-hoc testing revealed that only mCherry control mice differentially distributed their presses between the trained and the novel lever ($t_{(13)} = 5.63$, adjusted $p = 0.0002$), while the hM4D mice did not (adjusted $p = 0.15$). These results indicate that the OFC-M2 projection is functionally involved in learning to exploit known rules in an uncertain environment.

Figure 1.4. Chemogenetic attenuation of OFC-M2 projection neurons reduces exploitation of learned rules. Chemogenetic inhibition of OFC-M2 projection neurons throughout training and testing. **(a)** (top) Experimental timeline and (bottom) schematic of dual viral vector injection. **(b)** Representative images of mCherry and GFP fluorescence at 3.2x magnification in both OFC and M2. **(c)** Response rate during acquisition. mCherry = Fluorophore control mice expressing mCherry. hM4D = Inhibitory DREADD-expressing mice **(d)** Lever presses during outcome devaluation. There is a significant main effect of Valuation State. **(e)** Lever presses during novel lever test. RT = Random Time. CRF = Continuous Ratio of Reinforcement. RI = Random Interval. V/DV = Outcome Devaluation Test. Bars = ± SEM. n.s. = Not Significant, * = $p < 0.05$, ** = $p < 0.01$.

Figure 1.4


a
Experiment 4: OFC-M2 Inhibition
RT (1 day)
CRF (3 days)
RI30 (2 days)
RI60 (6 days)
V/DV Test (2 days)
Novel Lever Test (1 day)
AAV5/hSyn-DIO-hM4D-mCherry
OFC
M2
AAV5/CamKIIα-GFP-Cre
b
OFC
M2
mCherry
GFP
+2.7 Bregma
+0.5 Bregma
c
Acquisition Response Rate
Lever Presses/Min.
mCherry
hM4D
CRF
CRF
CRF
1
2
3
4
5
6
7
8
d
Devaluation Test
Lever Presses
mCherry
hM4D
Valued
Devalued
Valued
Devalued
e
Novel Lever Test
*
**
Lever Presses
mCherry
hM4D
Trained
Novel
Trained
Novel

Discussion

Our data suggests that exploitation and exploration are parallel decision processes, with OFC-M2 circuits supporting the acquisition and performance of exploitation. We have provided multiple, convergent lines of evidence indicating that mice do not generalize learned action contingencies in total during exploration on a novel lever, but instead choose to exploit known rules while they continue to explore for new rules associated with a novel lever. In support of this, learned outcome value of the trained lever does not appear to control novel lever pressing, nor does the amount of uncertainty experienced during learning. Instead, we find that experience with the learned rule predicts subsequent exploitation of that lever during testing, while that experience has little effect on continued exploration. In agreement with this, chemogenetic inhibition of OFC neurons projecting to M2 – a neural circuit involved in rule learning – was sufficient to induce greater exploration.

Attenuation of the OFC-M2 circuit revealed a functional role for this circuit in biasing exploitation of known rules. To our knowledge, this is the first time this circuit has been functionally manipulated whatsoever. OFC has a long history of research implicating it in representing outcome value (Stalnaker et al., 2015) and in reversal learning (Rudebeck & Murray, 2008), and has recently been proposed to incorporate expected uncertainty during decisions to guide behavior (Stolyarova & Izquierdo, 2017). A prominent hypothesis has been that the OFC represents the state space of a given task (Wilson et al., 2014). With regards to the latter hypothesis, an unanswered question is, where does OFC convey this state space information? OFC projections into amygdala (Fiuzat et al., 2017), and dorsal striatum (Gremel et al., 2016) appear to

convey information necessary for value-based decision-making, including broader state space representations (Stalnaker et al., 2016). Intracortical OFC projections have been largely neglected, but are interesting candidate regions for the conveyance of this state space information. One such cortical region is M2, which has been proposed to utilize evidence – both external sensory and internal information – to guide actions (Barthas & Kwan, 2017). What is unclear is whether M2 is directly computing and utilizing evidence, or whether this information arrives from other regions (Murakami et al., 2014, 2017). OFC is an interesting candidate source, given that M2 and OFC are reciprocally connected (Zingg et al., 2014), and bouton gain of OFC axons in M2 positively correlates with exploitation of learned rules, while bouton loss correlates with exploration (Johnson et al., 2016). This provides correlative evidence that OFC is indeed conveying task-relevant information to M2. Our results provide a causal link between activity in this pathway and subsequent decision-making, suggesting contribution of the OFC-M2 projection in arbitrating the exploitation of learned rules. Since we inhibited OFC-M2 projections throughout both training and test, we cannot conclude if this projection is also involved in using this learned information during novel lever testing. However, the results of the structural plasticity study (Johnson et al., 2016) would indicate that OFC-M2 projections are specifically involved in learning, particularly since there was no differences in structural plasticity between groups of mice that had to recall an already known rule vs. those that underwent a reversal.

We found no evidence for the involvement of OFC-M2 projections in goal-directed decision-making following outcome devaluation. This is somewhat surprising, as both OFC (Gremel & Costa, 2013b) and M2 (Gremel & Costa, 2013a) are

individually necessary for goal-directed control following outcome devaluation. In agreement with our current results, structural plasticity of OFC projection neurons in M2 does not correlate with the experience of reward alone, but instead specifically correlated with learning the relationship between actions and outcomes (Johnson et al., 2016). Thus it appears that while OFC projections into dorsal striatum (Gremel et al., 2016) and amygdala (Fiuzat et al., 2017) are involved in using value change to guide actions, we find no evidence that OFC projections to M2 convey outcome value; instead they may encode learned rules among outcomes, actions, and stimuli. Therefore, our findings suggest a projection-specific dissociation of OFC function, as we identify an OFC projection which may utilize state space representations provided by OFC to guide decision-making and action selection.

The results of our combined novel lever test and outcome devaluation study find no evidence that outcome value influences novel lever exploration. These results are significant on several different levels. Firstly, they replicate the finding that RR schedules bias goal-directed control over behavior and selective exploitation of the trained lever during a novel lever test, while RI schedules bias habitual control over behavior and exploration of the novel lever (Dickinson et al., 1983; Hilario et al., 2007, 2012). Thus, we were able to combine the novel lever test with the devaluation test and still replicate classical and long-standing schedule-induced differences in action control. This combination could prove useful, as it allows for the simultaneous study of different action control systems. This experiment also indicates that learned action-outcome associations do not generalize to the novel lever, as outcome value

manipulations – which control responding on the trained lever – have no effect on novel lever exploration.

It has been proposed that generalization on the novel lever test might occur as a result of a learned stimulus-response association generalizing to the perceptually similar novel lever (Hilario et al., 2012). However, we find that temporal uncertainty, which is known to increase habitual control over behavior (Derusso et al., 2010) has no effect on novel lever pressing. Additionally, we measured the duration of lever presses themselves (i.e., the response) in a separate experiment, and discovered that mice trained in RR schedules press the trained and novel lever differently. Thus, performance of the learned response itself does not completely generalize to the novel lever. It seems therefore that neither the stimulus-response relationship nor the response itself are fully generalized to the novel lever.

If novel lever pressing is not the sole result of generalization of learned rules, or of stimulus-response associations, what *is* controlling responding? It has recently been proposed that exploration is a distinct, early stage of learning which disappears following extended training (Iguchi et al., 2017). If exploration disappeared with training, we should expect a negative correlation between the amount of experience an animal had with the instrumental contingency and novel lever pressing. While we find evidence that the amount of experience correlates with trained lever pressing, there is no such relationship with novel lever pressing. Put another way, roughly the same level of novel lever exploration occurs regardless of the amount of experience animals have with the trained lever. Thus, animals might appear to explore early in training simply because there is relatively less exploitation occurring at this time point.

Classically, the explore/exploit dilemma is treated as a zero-sum game, where one necessarily excludes the other. While animals of course cannot simultaneously make explore/exploit-related actions, the trade-off between the two is not strictly zero sum as evidenced in the self-paced operant task used in this study. Mice in our task (and animals foraging in the wild) have many potential actions available to them – grooming, locomotion, making head entries – that do not explicitly fall into exploitation or exploration. It could be that the trial-based, forced choice structure of many tasks forces the apparent direct trade-off between exploration and exploitation. Our results suggest that the decision-making processes that arbitrate exploration and exploitation may not inherently be in competition; rather, they may run in parallel with action selection arising from the winning decision made (Ojeda et al., 2018). This is analogous to the current understanding of goal-directed and habitual action control systems as parallel processes, either of which may contribute to action control at a given time point (Balleine & O'Doherty, 2009). If exploration and exploitation decision processes do indeed run in parallel, an intriguing prediction is that it should be possible to selectively manipulate one or the other of these processes.

In support of this view, many studies have found different neuroanatomical substrates for exploration and exploitation (Addicott et al., 2017; Boorman et al., 2009; Daw et al., 2006; Laureiro-Martínez et al., 2014). However, other regions like the locus coeruleus (LC) have been implicated in both exploration and exploitation (Aston-Jones & Cohen, 2005). Interestingly, the LC is reciprocally connected with OFC (Aston-Jones & Cohen, 2005, p.) and M2 (Condé et al., 1995). It has been proposed that cortical input into LC is crucial for its ability to shift behavior between exploration and

exploitation (Aston-Jones & Cohen, 2005), and LC input into anterior cingulate (and adjacent M2) is critically involved in increasing behavioral variability that could underlie exploration (Tervo et al., 2014). LC norepinephrine is an important modulator of plasticity in the brain (Marzo et al., 2009); it is unknown if OFC-M2 projection plasticity might also be sculpted by LC norepinephrine input during learning.

We have provided evidence that novel exploration is unlikely to fully utilize previously learned rules about actions from the environment. This raises the possibility that the decision-making processes that arbitrate between exploration and exploitation may run in parallel and may not directly compete with one another.

Methods

Animals

Similar numbers of male and female C57BL/6J mice (> 7 weeks/50 PND) (The Jackson Laboratory, Bar Harbour, ME) were used for experiments. All procedures were conducted during the light period and mice had free access to water throughout the experiment. Mice were food restricted to 90% of their baseline weight 2 days prior to the start of experimental procedures, and were fed 1-4 hours after the daily training. All experiments were approved by the University of California San Diego Institutional Animal Care and Use Committee and were carried out in accordance with the National Institutes of Health (NIH) “Principles of Laboratory Care.” Mice were housed 2-4 per cage on a 14:10 light:dark cycle.

Acquisition

Mice were trained once per day in operant chambers in sound attenuating boxes (Med-Associates, St Albans, VT) in which they pressed a lever (left or right of the food magazine, counterbalanced for location) for an outcome of regular 'chow' pellets (20 mg pellet per reinforcer, Bio-Serv formula F0071). Each training session commenced with an illumination of the house light and lever extension and ended following schedule completion (30 reinforcers) or after 60-90 minutes had elapsed with the lever retracting and the house light turning off.

On the first day, mice were trained to approach the food magazine (no lever present) on a random time (RT) schedule, with a reinforcer delivered on average every 60 seconds for a total of 30 minutes. Next, mice were trained on a continuous ratio schedule of reinforcement (CRF) across 3 days, where every lever press was reinforced, with the total possible number of earned reinforcers increasing across days (CRF 5, 15, and 30).

Following CRF, mice were trained on either a random interval (RI) schedule to bias habitual control over actions (Dickinson, 1985) and action generalization (Hilario et al., 2012), or a random ratio schedule (RR) to bias goal-directed action control and action exploitation. In a RI(Y) schedule, the first lever press after an average of (Y) time has elapsed will be reinforced, using a probability distribution of $p = 0.10$ (e.g. in RI30, the first lever press after 30 seconds – on average – have elapsed will be rewarded). In a RR(X) schedule, on average (X) lever presses must occur before a reward is delivered. Initial training was conducted on a RI30 and RR10 for two days, followed by a progression to RI60 and RR20 (see each experiment for timeline details).

Generalization Testing

As described previously (Hilario et al., 2012), mice were placed in the training context and at session start two levers were extended; the previously trained lever as well as a novel, but identical lever in a different spatial location. Testing took place over 5 minutes and was conducted in the absence of reinforcement. Mice that made 0 presses on the trained lever were excluded from analyses.

Outcome Devaluation

Devaluation procedures occurred across two days. In brief, on the valued day, mice had ad libitum access to an outcome previously experienced in the home cage for 1 hour before being placed in the training context for a 5 minute, non-reinforced test session. On the devalued day, mice were given 1 hour of ad libitum access to the outcome previously earned by lever press, and then underwent a 5 minute, non-reinforced test session in the training context. The order of revaluation day was counterbalanced across mice. Mice who did not consume at least 0.1g of food on either the valued or devalued day were excluded.

Combined Outcome Devaluation and Generalization

Outcome devaluation was combined with the novel lever test such that both the trained and novel lever were presented following outcome devaluation via specific satiety. Testing occurred across two days, separated by one day of re-training in between. All conditions were counterbalanced between days.

Drugs

The hM4D-selective agonist Clozapine N-Oxide (CNO) was obtained from the National Institute of Mental Health (Bethesda, MD). The CNO dosage was 1.0 mg/kg at 10 ml/kg per mouse, delivered in saline via intraperitoneal injection. All mice were pretreated with CNO 30 minutes prior to the start of training or testing to allow for CNS penetration (Gremel & Costa, 2013b).

Surgical Procedure

For chemogenetic attenuation of OFC-M2, all viral vectors were obtained from the UNC Viral Vector Core (Chapel Hill, NC). Mice were anaesthetized with isoflurane (1-2%) and bilateral intracranial injections were performed via Hamilton (Reno, NV) syringe targeted at M2 (from Bregma: AP +0.5 mm, L ±0.5 mm and V -1.25 mm from the skull), or OFC (from Bregma: AP +2.7 mm, L ±1.65 mm and V -2.65 mm from the skull). Mice (n = 16) received 200nl of a viral vector (rAAV5/CamKIIα-GFP-Cre) expressing Cre recombinase (Cre) under the control of the calcium calmodulin dependent protein kinase II α (CamKIIα) in M2. In OFC, n = 8 mice received 200nl of a viral vector (rAAV5/hSyn-DIO-mcherry) as a control, and n = 8 mice received 200nl of a viral vector (rAAV5/hSyn-DIO-hM4D-mcherry) expressing a Cre-inducible, inhibitory DREADD (hM4D) coupled to a G_i signaling cascade which induces neuronal attenuation (Armbruster et al., 2007). Syringes were left in place for five minutes after injection to allow for diffusion. Mice were given at least two weeks of recovery before the start of experimental procedures. After behavioral testing was concluded, mice were euthanized and brains were extracted and fixed in 4% paraformaldehyde. The hM4D virus expressed the fluorescent marker mCherry, while the Cre virus expressed the fluorescent marker GFP. Localization and spread of viral expression was assessed

in 100 µm thick brain slices using fluorescent microscopy (Olympus MVX10). The final n's were: n =7 hM4D mice and n = 8 mCherry control mice.

Data Analysis

For all analyses, α = 0.05 was used as a threshold for significance. All analyses were two-tailed. Initial analyses were conducted to assess normal distributions and similar standard deviations. Where we found evidence for non-normal distributions or different standard deviations, we used Mann-Whitney tests. One-way or two-way repeated measures ANOVAs were used to examine acquisition and test data unless stated otherwise. The devaluation index was calculated by subtracting lever presses on the devalued day (DV) from lever presses on the valued day (V) and dividing by the total number of lever presses across both days (V - DV) / (V + DV). The generalization index was calculated by subtracting novel lever presses from trained lever presses and dividing by the total number of lever presses (Trained - Novel) / (Trained + Novel). Action-outcome contiguity was calculated by measuring the time in between a lever press and the next reinforcer delivery on average per animal. Behavioral data was recorded by MED-PC IV software, and analyzed in Excel, Matlab (Mathworks), Prism (Graphpad), and JASP.

Experiment 1: Role of outcome value in action generalization

16 C57BL/6J mice were used for this experiment. Two days prior to the start of behavioral procedures, mice were habituated to a novel cage for 1.5 hours which would later be used in the devaluation procedure. On schedule training days, mice were given a non-contingent, home cage outcome of 20% w/v sucrose (Sigma Aldrich,

St. Louis, MO) 1-4 hours after training, which would serve as a control for satiety during the devaluation test. Half of the subjects (n = 8) were trained under a RR schedule, while the other half (n = 8) were trained under a RI schedule of reinforcement. Mice were trained for 2 days on either RR10 or RI30, before being switched to a RR20 or RI60 schedule for 10 days of training prior to the combined outcome devaluation, action generalization test (Figure 1.1a). During the devaluation generalization test, several mice were excluded due to failing to consume the minimum during pre-feeding (0.1g either day), giving a final sample size n = 6 RI and n = 5 RR during the test.

Experiment 2: Role of uncertainty in action generalization

48 C57BL/6J mice were used for this experiment. Subjects were broken up into three different uncertainty groups using interval schedules of reinforcement, each with an initial n = 16. The three schedules used were a Fixed Interval (FI), a RI $p = 0.5$ and a RI $p = 0.1$ as described previously (Derusso et al., 2010). 3 mice were excluded for failing to acquire the task (1 from FI, 2 from RI $p = 0.5$) to give final sample sizes of n = 15 FI, n = 14 (RI $p = 0.5$), and n = 16 (RI $p = 0.1$). The schedules differed in terms of their reward probability distribution, but all shared the same average time to reward (Figure 1.2b). This was achieved by utilizing different time cycles (T) coupled with different probabilities (p). In the FI60 schedule, T = 60s and $p = 1.0$, such that at every 60s cycle, there is 100% chance of a reinforcer being earned following a lever press. In the RI60 ($p = 0.5$) schedule, T = 30s and $p = 0.5$, such that at every 30s cycle, there is a 50% chance of a press producing a reinforcer. In the RI60 ($p = 0.1$) schedule, T = 6s and $p = 0.1$, such that at every 6s cycle, there is a 10% chance of a press

producing a reinforcer. Mice were pre-trained on a RT and CRF schedule as described above, before being switched onto a FI30/RI30 schedule for 2 days, followed by 2 days of a FI60/RI60 schedule, then 1 day of novel lever testing, then 4 additional days of FI60/RI60 training, followed by a final day of novel lever testing (Figure 1.2a).

Experiment 3: Schedule-induced differences in action performance

7 C57BL/6J mice were used for this experiment, with n = 4 trained under a RI schedule and n = 3 trained under a RR schedule. During this experiment, the lever press durations were recorded. Mice were trained for 2 days on a RR10/RI30 schedule, followed by 10 days of RR20/RI60 training, followed by a novel lever test.

Experiment 4: Role of OFC to M2 projections in action generalization

16 C57BL/6J mice were used for this experiment. One hM4D mouse was excluded due to poor fluorophore expression leaving final n's at n = 8 mCherry controls and n = 7 hM4D mice. After pre-training, mice were trained for two days on a RI30 schedule, followed by 6 days of training on a RI60 schedule, followed by outcome devaluation testing. The following day, mice underwent a novel lever test (Figure 1.4a). CNO pretreatment began on the first day of schedule training and continued throughout training and testing.

Data Availability

The datasets generated and code used during the current study are available from the corresponding author on reasonable request.

Author Contributions

CG supervised the project. DS performed all the experiments and analyzed the data with input from CG. DS and CG conceptualized, wrote, and reviewed the manuscript.

Conflict of Interest

The authors declare no competing interests.

Acknowledgements

Chapter 1 is, in full, a reprint of the material published in *Scientific Reports* (2018). Schreiner, Drew C; Gremel, Christina M, 2018. The dissertation author was the primary investigator and author of this paper. This project was funded by the NIH: 4R00AA021780-02 (C.M.G), AA026077-01A1 (C.M.G.) and Whitehall Foundation (C.M.G).

References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A Primer on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research. *Neuropsychopharmacology*, *42*(10), 1931–1939. https://doi.org/10.1038/npp.2017.108

Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S., & Roth, B. L. (2007). Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(12), 5163–5168. https://doi.org/10.1073/pnas.0700293104

Aston-Jones, G., & Cohen, J. D. (2005). An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*, *66*(2), 315–326. https://doi.org/10.1016/j.neuron.2010.03.025

Balleine, B. W., & O'Doherty, J. P. (2009). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology*, *35*(1), 48–69. https://doi.org/10.1038/npp.2009.131

Barthas, F., & Kwan, A. C. (2017). Secondary Motor Cortex: Where 'Sensory' Meets 'Motor' in the Rodent Frontal Cortex. *Trends in Neurosciences*, *40*(3), 181–193. https://doi.org/10.1016/j.tins.2016.11.006

Beharelle, A. R., Polanía, R., Hare, T. A., & Ruff, C. C. (2015). Transcranial Stimulation over Frontopolar Cortex Elucidates the Choice Attributes and Neural Mechanisms Used to Resolve Exploration–Exploitation Trade-Offs. *Journal of Neuroscience*, *35*(43), 14544–14556. https://doi.org/10.1523/JNEUROSCI.2322-15.2015

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

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. *Neuron*, *62*(5), 733–743. https://doi.org/10.1016/j.neuron.2009.05.014

Bradfield, L. A., Dezfouli, A., van Holstein, M., Chieng, B., & Balleine, B. W. (2015). Medial Orbitofrontal Cortex Mediates Outcome Retrieval in Partially Observable

Task Situations. *Neuron*, *88*(6), 1268–1280. https://doi.org/10.1016/j.neuron.2015.10.044

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Condé, F., Maire-lepoivre, E., Audinat, E., & Crépel, F. (1995). Afferent connections of the medial frontal cortex of the rat. II. Cortical and subcortical afferents. *The Journal of Comparative Neurology*, *352*(4), 567–593. https://doi.org/10.1002/cne.903520407

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

Derusso, A. L., Fan, D., Gupta, J., Shelest, O., Costa, R. M., & Yin, H. H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. *Frontiers in Integrative Neuroscience*, *4*, 17. https://doi.org/10.3389/fnint.2010.00017

Dickinson, A. (1985). Actions and Habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *308*(1135), 67–78. https://doi.org/10.1098/rstb.1985.0010

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995). Motivational control after extended instrumental training. *Animal Learning & Behavior*, *23*(2), 197–206. https://doi.org/10.3758/BF03199935

Dickinson, A., Nicholas, D. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *35*(1), 35–51. https://doi.org/10.1080/14640748308400912

Fiuzat, E. C., Rhodes, S. E. V., & Murray, E. A. (2017). The Role of Orbitofrontal–Amygdala Interactions in Updating Action–Outcome Valuations in Macaques. *Journal of Neuroscience*, *37*(9), 2463–2470. https://doi.org/10.1523/JNEUROSCI.1839-16.2017

Gourley, S. L., Zimmermann, K. S., Allen, A. G., & Taylor, J. R. (2016). The Medial Orbitofrontal Cortex Regulates Sensitivity to Outcome Value. *The Journal of Neuroscience*, *36*(16), 4600–4613. https://doi.org/10.1523/JNEUROSCI.4253-15.2016

Gremel, C. M., Chancey, J. H., Atwood, B. K., Luo, G., Neve, R., Ramakrishnan, C., Deisseroth, K., Lovinger, D. M., & Costa, R. M. (2016). Endocannabinoid

Modulation of Orbitostriatal Circuits Gates Habit Formation. *Neuron*, *90*(6), 1312–1324. https://doi.org/10.1016/j.neuron.2016.04.043

Gremel, C. M., & Costa, R. M. (2013a). Premotor cortex is critical for goal-directed actions. *Frontiers in Computational Neuroscience*, *7*. https://doi.org/10.3389/fncom.2013.00110

Gremel, C. M., & Costa, R. M. (2013b). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3264

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*(7), 933–939. https://doi.org/10.1038/nn.2856

Hilario, M., Clouse, E., Yin, H. H., & Costa, R. M. (2007). Endocannabinoid Signaling is Critical for Habit Formation. *Frontiers in Integrative Neuroscience*, *1*. https://doi.org/10.3389/neuro.07.006.2007

Hilario, M., Holloway, T., Jin, X., & Costa, R. M. (2012). Different dorsal striatum circuits mediate action discrimination and action generalization: Neural circuits underlying action generalization. *European Journal of Neuroscience*, *35*(7), 1105–1114. https://doi.org/10.1111/j.1460-9568.2012.08073.x

Iguchi, Y., Lin, Z., Nishikawa, H., Minabe, Y., & Toda, S. (2017). Identification of an unconventional process of instrumental learning characteristically initiated with outcome devaluation-insensitivity and generalized action selection. *Scientific Reports*, *7*, 43307. https://doi.org/10.1038/srep43307

Johnson, C. M., Peckler, H., Tai, L.-H., & Wilbrecht, L. (2016). Rule learning enhances structural plasticity of long-range axons in frontal cortex. *Nature Communications*, *7*, 10785. https://doi.org/10.1038/ncomms10785

Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). The Nature of Belief-Directed Exploratory Choice in Human Decision-Making. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00398

Laureiro-Martínez, D., Canessa, N., Brusoni, S., Zollo, M., Hare, T., Alemanno, F., & Cappa, S. F. (2014). Frontopolar cortex and decision-making efficiency: Comparing brain activity of experts with different professional background during an exploration-exploitation task. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00927

Marzo, A., Bai, J., & Otani, S. (2009). Neuroplasticity Regulation by Noradrenaline in Mammalian Brain. *Current Neuropharmacology*, *7*(4), 286–295. https://doi.org/10.2174/157015909790031193

Morris, L. S., Baek, K., Kundu, P., Harrison, N. A., Frank, M. J., & Voon, V. (2016). Biases in the Explore–Exploit Tradeoff in Addictions: The Role of Avoidance of Uncertainty. *Neuropsychopharmacology*, *41*(4), 940–948. https://doi.org/10.1038/npp.2015.208

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

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

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A. M., Mortrud, M. T., Ouellette, B., Nguyen, T. N., Sorensen, S. A., Slaughterbeck, C. R., Wakeman, W., Li, Y., Feng, D., Ho, A., Nicholas, E., Hirokawa, K. E., Bohn, P., Joines, K. M., Peng, H., Hawrylycz, M. J., Phillips, J. W., Hohmann, J. G., Wohnoutka, P., Gerfen, C. R., Koch, C., Bernard, A., Dang, C., Jones, A. R., & Zeng, H. (2014). A mesoscale connectome of the mouse brain. *Nature*, *508*(7495), 207–214. https://doi.org/10.1038/nature13186

Ojeda, A., Murphy, R. A., & Kacelnik, A. (2018). Paradoxical choice in rats: Subjective valuation and mechanism of choice. *Behavioural Processes*, *152*, 73–80. https://doi.org/10.1016/j.beproc.2018.03.024

Ostlund, S. B., Winterbauer, N. E., & Balleine, B. W. (2009). Evidence of Action Sequence Chunking in Goal-Directed Instrumental Conditioning and Its Dependence on the Dorsomedial Prefrontal Cortex. *The Journal of Neuroscience*, *29*(25), 8280–8287. https://doi.org/10.1523/JNEUROSCI.1176-09.2009

Rhodes, S. E. V., & Murray, E. A. (2013). Differential Effects of Amygdala, Orbital Prefrontal Cortex, and Prelimbic Cortex Lesions on Goal-Directed Behavior in Rhesus Macaques. *Journal of Neuroscience*, *33*(8), 3380–3389. https://doi.org/10.1523/JNEUROSCI.4374-12.2013

Rothermel, M., Brunert, D., Zabawa, C., Díaz-Quesada, M., & Wachowiak, M. (2013). Transgene Expression in Target-Defined Neuron Populations Mediated by Retrograde Infection with Adeno-Associated Viral Vectors. *Journal of Neuroscience*, *33*(38), 15195–15206. https://doi.org/10.1523/JNEUROSCI.1618-13.2013

Rudebeck, P. H., & Murray, E. A. (2008). Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *The Journal of*

Neuroscience : The Official Journal of the Society for Neuroscience, *28*(33), 8338–8343. https://doi.org/10.1523/JNEUROSCI.2272-08.2008

Shan, Q., Ge, M., Christie, M. J., & Balleine, B. W. (2014). The Acquisition of Goal-Directed Actions Generates Opposing Plasticity in Direct and Indirect Pathways in Dorsomedial Striatum. *Journal of Neuroscience*, *34*(28), 9196–9201. https://doi.org/10.1523/JNEUROSCI.0313-14.2014

Siniscalchi, M. J., Phoumthipphavong, V., Ali, F., Lozano, M., & Kwan, A. C. (2016). Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior. *Nature Neuroscience*, *19*(9), 1234–1242. https://doi.org/10.1038/nn.4342

Stalnaker, T. A., Berg, B., Aujla, N., & Schoenbaum, G. (2016). Cholinergic Interneurons Use Orbitofrontal Input to Track Beliefs about Current State. *The Journal of Neuroscience*, *36*(23), 6242–6257. https://doi.org/10.1523/JNEUROSCI.0157-16.2016

Stalnaker, T. A., Cooch, N. K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. *Nature Neuroscience*, *18*(5), 620–627. https://doi.org/10.1038/nn.3982

Stolyarova, A., & Izquierdo, A. (2017). Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning under uncertainty. *ELife*, *6*, e27483. https://doi.org/10.7554/eLife.27483

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. 551.

Tervo, D. G. R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., & Karpova, A. Y. (2014). Behavioral Variability through Stochastic Choice and Its Gating by Anterior Cingulate Cortex. *Cell*, *159*(1), 21–32. https://doi.org/10.1016/j.cell.2014.08.037

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953–956. https://doi.org/10.1038/35082081

White, I. M., & Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Experimental Brain Research*, *126*(3), 315–335.

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

Yin, H. H. (2009). The role of the murine motor cortex in action duration and order. *Frontiers in Integrative Neuroscience*, *3*, 23. https://doi.org/10.3389/neuro.07.023.2009

Zingg, B., Hintiryan, H., Gou, L., Song, M. Y., Bay, M., Bienkowski, M. S., Foster, N. N., Yamashita, S., Bowman, I., Toga, A. W., & Dong, H.-W. (2014). Neural Networks of the Mouse Neocortex. *Cell*, *156*(5), 1096–1111. https://doi.org/10.1016/j.cell.2014.02.023

CHAPTER TWO

Mice are not automatons; subjective experience in premotor circuits guides behavior

Drew C. Schreiner[1], Christian Cazares[2], Rafael Renteria[1], and Christina M. Gremel[1,2]

[1]Department of Psychology, University of California San Diego, La Jolla, CA 92093

[2]The Neurosciences Graduate Program, University of California San Diego, La Jolla, CA 92093

Abstract

Subjective experience is a powerful driver of decision-making and continuously accrues. However, most neurobiological studies constrain analyses to task-related variables and ignore how continuously and individually experienced internal, temporal, and contextual factors influence adaptive behavior during decision-making and the associated neural mechanisms. We show mice rely on learned information about recent and longer-term subjective experience of variables above and beyond prior actions and reward, including checking behavior and the passage of time, to guide self-initiated, self-paced, and self-generated actions. These experiential variables were represented in secondary motor cortex (M2) activity and its projections into dorsal medial striatum (DMS). M2 integrated this information to bias strategy-level decision-making, and DMS projections used specific aspects of this recent experience to plan upcoming actions. This suggests diverse aspects of experience drive decision-making and its neural representation, and shows premotor corticostriatal circuits are crucial for using selective aspects of experiential information to guide adaptive behavior.

Keywords: subjective experience, decision-making, secondary motor cortex, striatum, corticostriatal

Introduction

Most neurobiological investigations into decision-making seek to use well-constrained tasks to isolate specific components of decision-making and illuminate the corresponding neural mechanisms. These investigations often institute a trial structure, limit choice and movement, and elicit behavior via cues, with the latter leading to a historical focus on elicited stimulus-response characterization of involved mechanisms (Juavinett et al., 2018). Thus, interpretation of the associated mechanisms and models are made within this well-constrained vacuum. There is growing concern that such an approach negates the very individualistic, experiential, and continuous nature of decision-making (Balleine, 2019; Gomez-Marin et al., 2014; Krakauer et al., 2017; Schreiner et al., 2021; Yoo et al., 2021). Presumably, the continuous experiential information accrued by the self is reflected in and used by the brain to execute adaptive behavior to support ongoing decision-making. Yet such information is often treated as task-irrelevant and ignored or factored out (Roy et al., 2021), to an at best incomplete, or at worst inaccurate picture of involved neural mechanisms. Indeed, the seemingly widespread distribution of similar decision-making information across the brain (Allen et al., 2017; Steinmetz et al., 2019) may in part be due to a lack of accounting for the unique constellation of internal, experiential, temporal, and contextual information encountered by an individual that drives decision-making, referred to here as “subjective experience”.

Evidence from ethological approaches shows that diverse types of subjective experience contribute to decision-making. When dropping shelled prey to break them, a crow will integrate the type of prey, the number of times the item has been dropped,

the hardness of the surface, and the amount of kleptoparasitism to determine how high to drop the item (Cristol & Switzer, 1999). Behavior need not be complex; even innate behaviors are modified by experience (Remedios et al., 2017) and simple experience-based strategies such as win-stay and lose-shift persist after performance (a proxy for learning) plateaus. Experiential-based emergence of control does not happen within a vacuum; an individual's interactions with their environment drive adaptive behavior (Balleine, 2019; Costa, 2011) with contributions from temporal (Ariely & Zakay, 2001), historical and contextual (Bouton & Balleine, 2019), and internal state factors (Balleine & Dickinson, 1998; Berridge et al., 2008). Indeed, there is an interplay between exploration and the accrual of experiential information, with exploration uncovering contingency and consequence information that in turn can be used to bias towards further exploration or experience-guided decision-making. However, in many investigations, experience is either not needed or even actively detrimental to performance (e.g., in perceptual decision-making tasks, subjects should ideally attend only to the current stimulus). That subjective experience appears to contribute even where it is "unnecessary" argues that it is a powerful driver of adaptive behavior (Lak et al., 2020). Further, many investigations constrain behavior by using trial-based, binary decision tasks unlikely to generalize fully to more ethological, self-generated, and continuously evolving types of decision-making (Yoo et al., 2021).

As subjective experience can play a large role in psychiatric disease (e.g., the temporal pattern of drug use is decisive in substance use disorders (Allain et al., 2015)), there is a need to account for its influence on the behavioral and neural mechanisms of decision-making. One neural circuit that is disrupted in disease and

presents as a candidate for the integration of such experiential information is secondary motor cortex (M2) (Ebbesen et al., 2018). On one hand, M2's sensory, motor, and premotor characteristics have implicated a role in using experience to guide decision-making (Erlich et al., 2011; Murakami et al., 2014; Pinto et al., 2019; Siniscalchi et al., 2016). On the other hand, several studies have found that M2 appears to be involved in implementing stochastic or exploratory decisions (Murakami et al., 2017; Pisupati et al., 2021; Schreiner & Gremel, 2018; Tervo et al., 2014). However, animals may decide to explore *based upon their experience*; for instance making more exploratory decisions when uncertainty is high (Dhawale et al., 2019). Thus, attribution of M2 function to seemingly disparate processes may reflect the lack of accounting for or limiting the contribution of experiential information. Instead, we hypothesize that M2 represents and integrates experiential information to guide experience or exploration-based decision-making when use of such information is advantageous. This strategy-level control over action selection may be exerted through M2 projections into dorsal medial striatum (DMS) (Delevich et al., 2020; Hintiryan et al., 2016). Indeed, recently M2-DMS projections have been implicated in repetitive actions in a mouse model of Obsessive Compulsive Disorder (Corbit et al., 2019), suggesting that M2-DMS dysfunction may contribute to disease states. In order to capture the contribution of experiential information to decision-making and its potential neural implementation, we need to investigate within a framework where the use of experience is *essential* and not merely incidental to decision-making. Here, we utilized an unstructured free operant foraging task with continuous variables in mice where experience is crucial for efficient performance. We find aspects of experiential

information normally considered task-irrelevant play large roles in supporting adaptive behavior, often more than that played by reward itself. We then show M2 circuits and their output to dorsal medial striatum (DMS) are crucial for use of this subjective experience to drive adaptive behavior.

Results

Mice learned an unstructured, self-generated, self-paced lever press hold down task

We adapted an instrumental task (Fan et al., 2012; Platt et al., 1973; Yin, 2009) where mice (n = 12 C57BL/6J) were trained to press and hold down a lever for at least a minimum duration to earn a food reward, with reward delivered at lever press release/offset (Figure 2.1A). There were no external cues signaling reward availability or duration, nor any trial structure (lever was always available). Thus lever presses were self-initiated, self-paced, and self-terminated and mice had to explore the contingency duration to determine the rule, a process termed action differentiation (Skinner, 1938).

We first examined macroscopic aspects of lever pressing. Mice were initially trained with an >800ms criterion before being shifted to a >1600ms criterion. Mice readily learned that press duration was the operant and quickly reduced the number of Total Lever Presses (Figure 2.1B; 1-way ANOVA $F_{2.9, 31.9} = 12.0$, $p < 0.0001$), while they increased the percentage of presses that met the minimum duration criterion (Figure 2.1C; referred to as %Presses Met Criteria, 1-way ANOVA $F_{4.22, 46.5} = 17.2$, $p < 0.0001$), and showed little evidence of stereotypies in their lever pressing (see Note

S2.1). Mice were sensitive to the minimum duration rule and shifted the distribution of press durations from a pretraining session with no duration requirement, to the final day of >800ms training, and further still to the final day of >1600ms training (Figure 2.1D; 2-way RM ANOVA, main effect of Duration Bin $F_{31,1056} = 34.1$, $p < 0.0001$, and an interaction (Duration Bin/Criterion) $F_{62,1056} = 10.5$, $p < 0.0001$). To examine whether actions were controlled by their expected consequence and operationally goal-directed or were instead habitual (Adams & Dickinson, 1981; Dickinson, 1985), we performed outcome devaluation testing (Figure 2.S1A). Mice reduced their Total Lever Presses on Devalued days relative to Valued days (Figure 2.S1B), consistent with using expected outcome value to guide decisions as seen in goal-directed control (Adams & Dickinson, 1981). Although Total Lever Presses decreased, the %Presses Met Criteria *increased* following devaluation (Figure 2.S1C) with a small rightward shift in the distribution of press durations (Figure 2.S1D), suggesting action selection and execution may be differentially controlled by outcome value.

It is clear that mice can use contingency and consequence information to perform this task, but it is unclear *how* they are doing so. One possibility is that executed lever press durations are independent, with mice timing each press. If so, we hypothesized that mice may exhibit the scalar property of timing; as lever press durations increase, so too does variability (Gibbon et al., 1984; Yin, 2009). We calculated the median and interquartile range (IQR) of each animal's lever press durations across training (Figure 2.1E) and found concomitant increases in both the median and the IQR across training days during initial short criterion training (2-way RM ANOVA, main effect only of Day ($F_{5,55} = 19.5$, $p < 0.0001$). However, when the

duration criterion increased and training continued, the pattern of change in lever press IQR departed from the pattern of change in lever press median duration (2-way RM ANOVA, main effect of Day $F_{7,77} = 14.0$, $p < 0.0001$, and an interaction (Median/IQR x Day) $F_{7,77} = 2.44$, $p = 0.026$), suggesting reduced reliance on timing information. Reminiscent of skill learning, mice also showed within session increases in median durations during both the first and *last* day of training (Figure 2.1F). Linear regressions showed a significantly non-zero slope on both the first >800ms day ($F_{1,110} = 28.9$, $p < 0.0001$, $R^2 = 0.21$) as well as the final >1600ms day ($F_{1,115} = 12.6$, $p = 0.0006$, $R^2 = 0.099$). However, although within session increases in IQR were present on the first day of training ($F_{1,110} = 48.5$, $p < 0.0001$, $R^2 = 0.306$), by the final day of training IQR no longer increased within a session ($F_{1,115} = 0.28$, $p = 0.59$, $R^2 = 0.002$). Furthermore, while the slopes of the within session median and IQR were not different on the first day of training ($p = 0.27$), they *were* different by the final day ($F_{1,230} = 9.1$, $p = 0.003$), in violation of the scalar property of timing. This suggests mice used additional information other than solely timing behavior to control lever pressing.

What is this non-timing information? One possibility is that mice relied on recent experience to guide their decision-making. In Figure 2.1G, we plot press durations across one session for one well-trained mouse. Immediately clear is the rich behavior, both in terms of when presses occurred, and in their duration. However, there also appeared to be distinct periods of reduced variability. A cumulative sum (upper bound) analysis (Figure 2.1H) uncovered prolonged periods of time when mice emitted press durations >2 standard errors (SE) above the mean, (Figures 2.1I-J). This was not due to random chance, or the consequence of very long press durations inflating the

cumulative sum (2-way RM ANOVA; percentage of >2SE presses: main effect of Day $F_{3,33} = 3.98$, $p = 0.016$, main effect of Actual vs. Shuffled $F_{1,11} = 17.1$, $p = 0.0017$. Consecutive >2SE presses: main effect only of Actual vs. Shuffled $F_{1,11} = 14.0$, $p = 0.0032$).

Figure 2.1. *Mice learned an unstructured, self-generated, self-paced lever press hold down task.* **(A)** Behavioral schematic; mice learn to press and hold down a lever for at least a minimum duration to earn food reward. **(B)** Total Lever Presses across training days. **(C)** %Presses that met criteria across training. **(D)** Histogram of lever press durations (100ms bins) on the final pretraining day (CRF = Continuous Ratio of Reinforcement), and the final 800ms and 1600ms days. Dashed lines indicate criterion. **(E)** Median and Interquartile Range (IQR) of lever press durations across training days. **(F)** Duration median (Med) and IQR within a session, grouped by cumulative number of rewards. **(G)** Sample behavior of one trained mouse showing press durations in order of occurrence. **(H)** Upper cumulative sum from the same mouse/session. **(I)** Number of consecutive presses and **(J)** Overall % of presses that were >2 Standard Errors (SE) above the mean in the upper cumulative sum. 800ms and 1600ms refer to days where criterion was >800ms or >1600ms. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < .05$. Points represent mean+SEM across mice, unless noted otherwise.

Figure 2.1

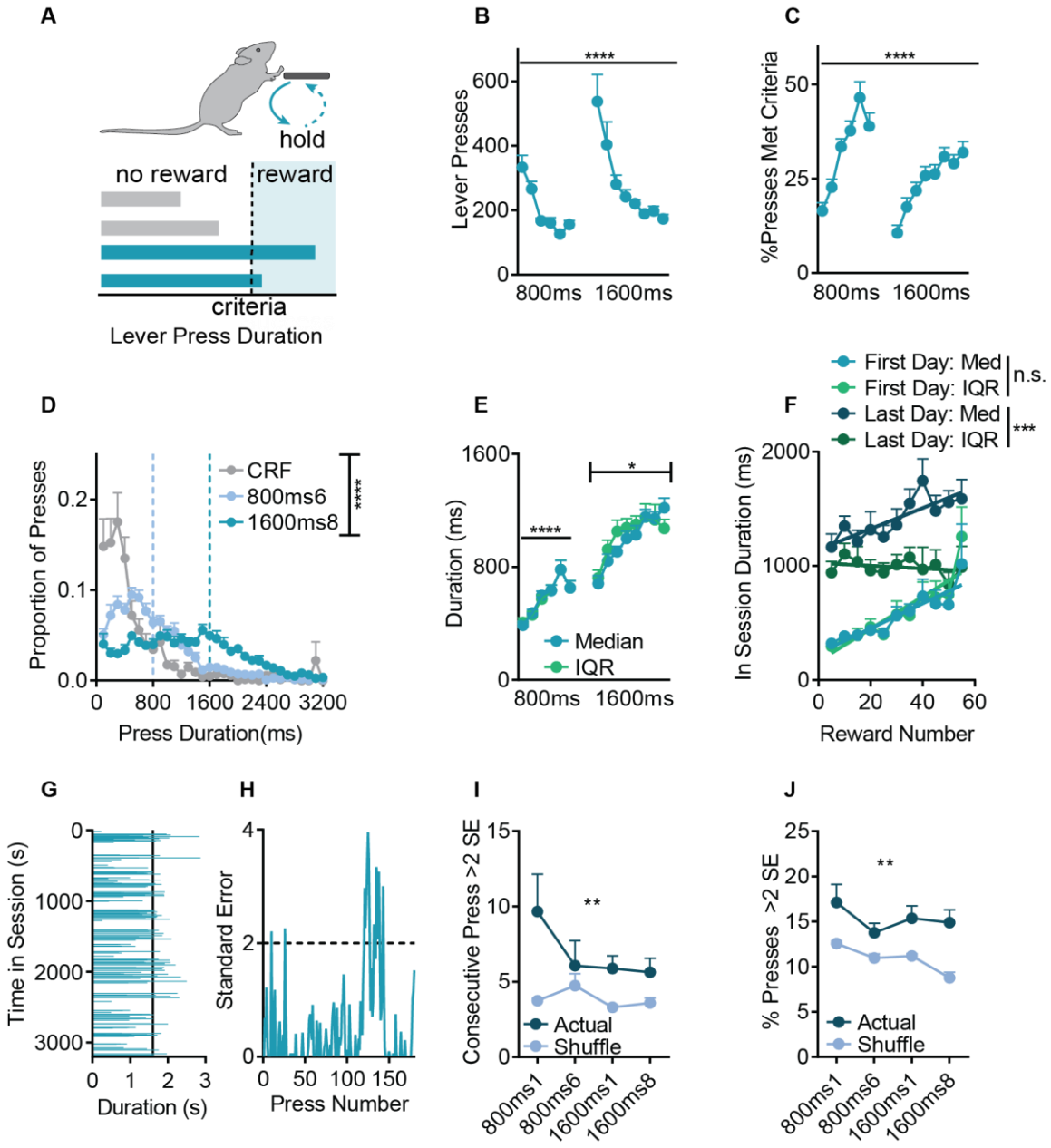

A
hold
no reward
reward
criteria
Lever Press Duration
B
****
Lever Presses
600
400
200
0
800ms 1600ms
C
****
%Presses Met Criteria
50
25
0
800ms 1600ms
D
CRF
800ms6
1600ms8
****
Proportion of Presses
0.2
0.1
0.0
0 800 1600 2400 3200
Press Duration(ms)
E
*
****
Duration (ms)
1600
800
0
Median
IQR
800ms 1600ms
F
First Day: Med
First Day: IQR
n.s.
Last Day: Med
Last Day: IQR
***
In Session Duration (ms)
2000
1000
0
0 20 40 60
Reward Number
G
Time in Session (s)
0
1000
2000
3000
0 1 2 3
Duration (s)
H
Standard Error
4
2
0
0 50 100 150
Press Number
I
**
Consecutive Press >2 SE
15
10
5
0
Actual
Shuffle
800ms1
800ms6
1600ms1
1600ms8
J
**
% Presses >2 SE
25
20
15
10
5
0
Actual
Shuffle
800ms1
800ms6
1600ms1
1600ms8

Figure 2.S1

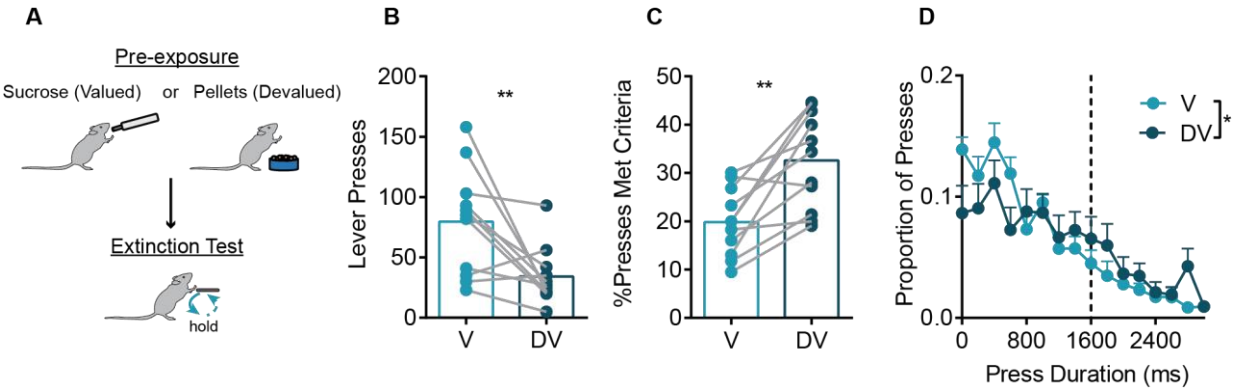


Figure 2.S1, related to Figure 2.1. *Outcome value has differential control over action selection and execution.* **(A)** Schematic of the outcome devaluation procedure. In a within subjects design, mice received 1 hr of pre-exposure to either sucrose (Valued, V) or pellets (Devalued, DV), followed by a 5 minute extinction test across 2 days (order counterbalanced). **(B)** Total Lever Presses on V and DV days. Paired t-test, $t_{10} = 3.09$, $p = 0.012$. **(C)** %Presses that met criteria on V and DV days. Paired t-test, $t_{10} = 4.55$, $p = 0.0011$. **(D)** Histogram of press durations on V and DV days (200ms bins). 2-way RM ANOVA, main effect of Duration Bin, $F_{15,150} = 12.1$, $p < 0.0001$ and an interaction (Duration Bin x V/DV) $F_{15,150} = 2.19$, $p = 0.009$. Data in (D) are mean+SEM across mice, bars in (B-C) are mean. ** $p < 0.01$, * $p < 0.05$.

Subjective experience contributes to internally-generated decision-making

The relative similarity among serial lever presses suggests that recent experience may contribute to adaptive behavior in support of self-generated decision-making. We modeled the effect of recent press history on performance by creating a simple linear mixed effect model (LME) that sought to predict current press duration (n) given recently executed durations (n-back). We included random effects of both training day and mouse to account for the repeated structure of our data. We also included several control variables and compared the actual coefficients to those obtained from order shuffled data using permutation tests (Table 2.S1). We found a consistent significant linear relationship between current press n duration and the durations of n - 1 through n - 6 presses, with the magnitude of this relationship decaying across n-back presses (Figure 2.2B). This suggests that recent subjective experience contributes to continuous decision-making.

However, recent lever presses are not the only experiential information available (Figure 2.1A). The unstructured nature of this self-generated task allows us to capture aspects of decision-making that occur across a continuous space beyond just the press duration itself. We created more complex LMEs, first building a “full” model that included n - 1 through n - 6 durations, as well as main effect and interaction terms for other n-back variables, such as the inter-press interval between press n and press n - 1 (see Table 2.S2 for terms). We performed backwards selection on this full model using Bayesian Information Criterion (BIC), leaving us with the model in Table 2.S3. Follow-up permutation tests found that all the variables identified by BIC

selection also significantly differed from order shuffled data (shuffled within a single mouse/session), suggesting that it is indeed the experienced order of these variables that affects subsequent press duration and not correlation across mice or across days. Importantly, when we built multiple linear regressions with the same experiential predictors, but using only individual mouse/session data, we found a strong correlation between model R^2 and task performance (repeated measures correlation, $R_m = 0.56$, DF =153, $p < 0.0001$, slope = 0.38; Figure 2.S2A). Thus, use of experience (as these models use only experience to predict duration) correlated with performance. Additionally, in the combined (all mice and training days) LMEs, there was a significant positive interaction between performance and the use of recent experience (serially adjacent presses were more similar in mice with higher performance, Figure 2.2D shows that an increase in %Met Criteria of 30% would roughly double the n/n - 1 relationship).

Of note, our goal with this model was not to make the most accurate predictions (though it predicted 24.1% of all lever press durations within a 95% CI, and accurately predicted whether a press did or did not meet criteria 73.8% of the time). Instead, we sought to ascertain 1) which experiential variables contributed, 2) if these variables interacted with previously made lever press durations, and 3) whether recent (i.e., n - 1) versus longer-term (i.e., a moving average of durations from n - 7 through n - 60) lever press duration experience differentially contributed (Iigaya et al., 2018). Beginning with the latter, we found that this long-term duration moving average coefficient significantly differed from order shuffled data (Figure 2.2C; permutation test

$p < 0.001$). Thus, both recent and longer-term duration experience contributed to emitted durations.

Observational, but not reward information contributes to decision-making

Experience can also be driven by information feedback processes. Here, headentry into the food magazine (HE) as a checking behavior provided information as to whether a lever press was or was not successful. HE behavior increased the relationship both between press n and n - 1, and between press n and the moving average (Figure 2.2D). The magnitude of this increase was quite large: lever presses within a lever press/HE/lever press sequence were effectively twice as related to one another relative to those in a lever press/lever press sequence. Thus checking behavior was a source of experiential information and influenced the subsequent executed behavior.

Success feedback may also be signaled by reward, and reward feedback is a crucial aspect of many decision-making and learning theories (Rescorla & Wagner, 1972; Sutton & Barto, 1998). Reward can modulate win-stay and lose-shift strategies even in well-learned tasks (Busse et al., 2011; Lak et al., 2020). However, it is less clear how reward might modulate decision-making in a more unconstrained task where one dictates their own behavioral opportunities. In regard to the present task, if mice “won and stayed”, we should expect that earning a reward on press n - 1 would cause mice to make a similar duration press afterwards.

We found no evidence of simple win-stay/lose-switch behavior, either in the use of recent (n - 1), or long-term (moving average) experience (Figure 2.2D). Put another way, earning a reward did not increase the similarity between press n and press n - 1, nor did failing to earn a reward lead to drastic shifts in behavior. We reasoned that perhaps the lack of a win-stay effect may have been due to reward being deterministic, and thus, only errors of execution could occur (McDougle et al., 2019). Therefore, we imposed a probabilistic reward schedule in a separate cohort (25%, 50%, or 75% rewarded, n = 5 mice per group) following training. Here, the %Met Criteria increased (Figure 2.S2B). A Met Criteria press - whether or not it was rewarded - led to an increased relationship between press n and n - 1 (significant positive interaction, Figure 2.S2C). The magnitude of this effect was *larger* when the Met press was unrewarded, and this “win-stay” effect was more pronounced in the groups where a Met press was least likely to produce a reward. Thus, mice made a press that was more similar to the one that preceded it after a Met press, especially if that press happened to be unrewarded due to chance. This provides evidence mice used an internal representation of press duration to guide behavior and relied less on the presence of reward.

Time contributes to and modifies use of subjective experience in decision-making

The above findings challenge the assumption that decision-making is solely determined by the serial order of actions and their outcome, as is often presumed in trial-based experimental designs. That sources of this crucial experiential information, such as checking behaviors, accrue across a continuous temporal space raises the

question of how the passage of time itself may influence decision-making. We find the relationship between two adjacent presses (presses n and n - 1) decreased as the inter-press-interval (IPI) increases (Figure 2.2E). To give an example of the magnitude, the model predicts that the relationship between n and n - 1 would be approximately 0 if they are separated by 120s. This raises the hypothesis that animals may rely more on the long-term moving average to guide their behavior following long IPIs (Iigaya et al., 2018). Indeed, the use of long-term experience was unaffected by the IPI. Further, we found that n and n - 1 became more similar towards the end of a session, and again, there was no relationship between time in session and the moving average (Figure 2.2E). Collectively, these results suggest that the passage of time is a crucial aspect to modifying recent experience, with less effect on the contribution of long-term learned contingencies.

Figure 2.2

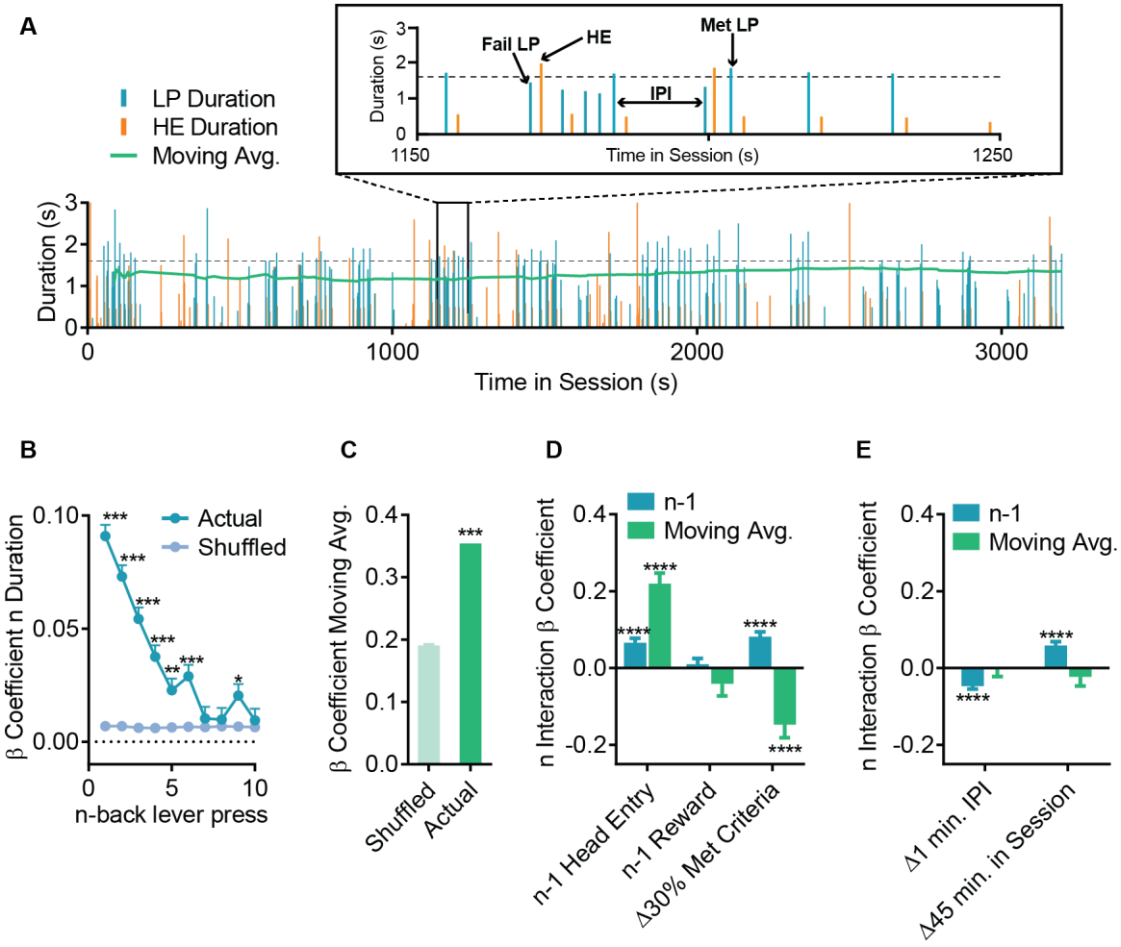


Figure 2.2. *Subjective experience contributes to internally-generated decision-making*. **(A)** Sample data from one mouse (as in Figures 2.1G-H) showing the diversity of experiential information available. Top shows a zoomed in subset. Dashed line indicates 1600ms criterion. **(B)** β coefficients of LME model relating current lever press duration (n) to preceding press durations (n - x) for Actual and order Shuffled data. **(C)** Moving average β coefficient for Actual and Shuffled data. **(D-E)** β coefficients for the interaction between experiential variables and recent (n - 1 duration) or long-term (moving average) experience. For display purposes we transformed continuous variables to show relevant changes, e.g. time in session, which is in units of ms, was transformed to 45 min or half the duration of a session. LP = Lever Press, HE = Headentry into food magazine, IPI = Inter Press Interval. Δ = Change. * Markers in D-E indicate significant F-tests on model terms. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < .05$. Data points are mean+SEM. Shuffled data are mean+SEM of 1000 order shuffled β coefficients. See also Figure 2.S2 and Tables 2.S1-3.

Figure 2.S2

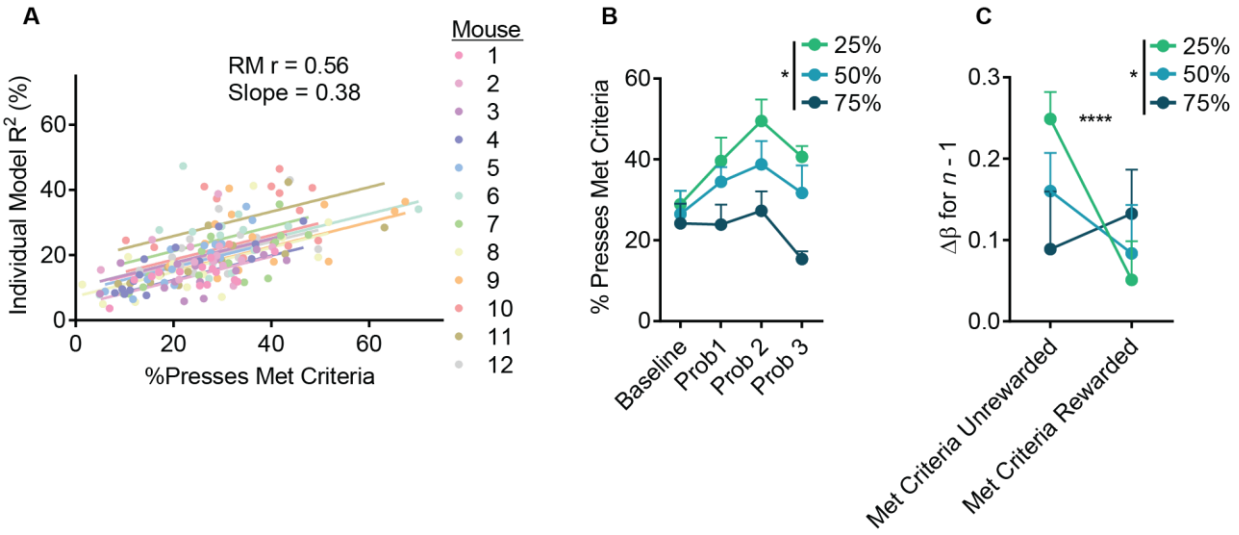


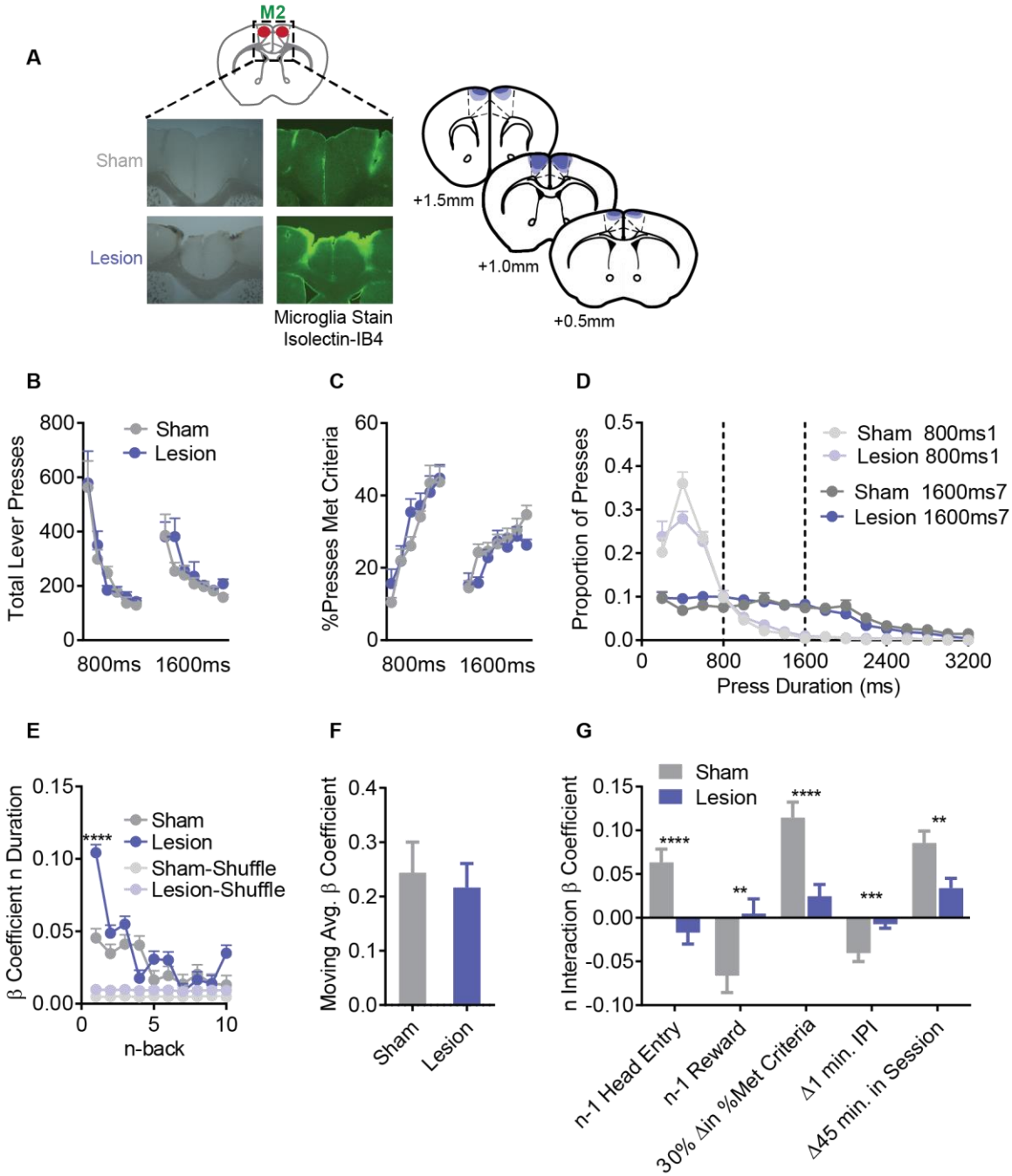
Figure 2.S2, related to Figure 2.2. *Probabilistic reward induces win-stay behavior and use of experience correlates with task performance.* **(A)** Repeated measures correlation between task performance (%Presses Met Criteria) and model fit (R^2) when building LMEs using individual mouse/session data. Intercept was allowed to vary across mice while using a common slope. **(B-C)** Following initial training on 100% reward, mice were shifted to either 25%, 50%, or 75% reward and trained for 3 days. **(B)** %Presses met criteria across training. 2-way ANOVA (Probability x Day), no interaction, main effects of Day $F_{3,36} = 6.34$, $p = 0.0015$, and probability group $F_{2,12} = 5.28$, $p = 0.0226$. **(C)** β coefficients for the interaction between presses that met criteria and n - 1 duration. Interaction between n - 1 duration and n - 1 outcome (Met Yes Reward vs. Met No Reward): $F_{2,17403} = 30.2$; $p < 0.0001$) and 3-way interaction between Duration, Outcome, and Group: $F_{4,17403} = 2.59$; $p = 0.035$. Baseline = Final pretraining day with 100% reward. Prob 1, 2, 3 = Training day 1, 2, or 3 of probability training. RM r = Repeated Measures correlation R. **** $p < 0.0001$, * $p < 0.05$.

M2 represents prior experience and may guide exploration

M2 has been reported to be involved in both exploration and experience-based decision-making, and this apparent discrepancy may be due to neglecting the contribution of some of these checking, contextual, and temporal variables. Therefore, we performed pretraining lesions of M2 using ibotenic acid (Figure 2.3A; Lesion n = 10) or vehicle (Sham n = 8). In line with prior reports (Yin, 2009), we found no differences between Sham and Lesion mice in coarse behavioral measurements such as Total Lever Presses (Figure 2.3B), %Presses Met Criteria (Figure 2.3C), or Press Durations (Figure 2.3D). However, M2 lesioned mice executed lever press durations that were more similar to their prior action (Figure 2.3E). This was evidenced by a specific increase in the magnitude of the n - 1 β coefficient compared to Sham mice (2-way ANOVA (n-back x Sham/Lesion) main effect of n-back ($F_{9,572740} = 20.2$, $p < 0.0001$) and Sham/Lesion ($F_{1,572740} = 14.1$, $p = 0.0002$) and significant interaction ($F_{9,572740} = 6.33$, $p < 0.0001$). Post-hoc testing revealed a significant group difference only at n - 1 ($t_{572740} = 6.87$, $p < 0.0001$), with no differences at further n-back presses, nor in the moving average term (Figure 2.3F). Using the complex LME model, M2 lesions disrupted all n - 1 duration interactions, including Reward, Checking, IPI, and Time in Session (Figure 2.3G, see also Table 2.S4). Lesions did not affect moving average interactions. Collectively, this suggests M2 lesioned mice were relatively inflexible, akin to previous studies where M2 lesions biased use of habitual or model-free processes (Gremel & Costa, 2013), and were left to rely on the just-made action without integration of broad experiential information.

Figure 2.3. *Pretraining lesions of M2 impair integration of experiential information.* **(A)** (top left) Schematic, (bottom left) sample histology, and (right) average/maximal (dark/light shading) spread of excitotoxic lesions of M2, slice coordinates relative to Bregma. **(B)** Total Lever Presses across training days. **(C)** %Presses that met criteria across training days. **(D)** Histogram of lever press durations on the first and last days of the 800ms and 1600ms criteria (200ms bins). **(E)** β coefficient from LME models predicting n duration from n-back durations for Actual and order Shuffled data. **(F)** β coefficient for the moving average term. **(G)** β coefficients for the interaction terms from the complex LME model. For display purposes we transformed the continuous variables to show relevant changes. B-D are mean+SEM across mice. Shuffled data are the mean+SEM of 1000 order shuffled β coefficients. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$. See also Table 2.S4.

Figure 2.3


A
M2
Sham
Lesion
Microglia Stain
Isolectin-IB4
+1.5mm
+1.0mm
+0.5mm
B
Total Lever Presses
Sham
Lesion
800ms
1600ms
C
%Presses Met Criteria
D
Proportion of Presses
Sham 800ms1
Lesion 800ms1
Sham 1600ms7
Lesion 1600ms7
Press Duration (ms)
E
β Coefficient n Duration
Sham-Shuffle
Lesion-Shuffle
n-back
F
Moving Avg. β Coefficient
G
n Interaction β Coefficient
n-1 Head Entry
n-1 Reward
30% Δin %Met Criteria
Δ1 min. IPI
Δ45 min. in Session

How may subjective experience influence representation of decision-making in M2 circuits? We utilized *in vivo* fiber photometry (Figure 2.4A; coordinates from Bregma: AP +1.0mm, L ±0.5mm and V −1.2mm, -1.4mm from the skull), and measured population Ca^{2+} activity from M2 excitatory neurons (n = 8 mice). Aligning baseline z-scored activity to lever press onset (Figure 2.4B), we observed preceding ramping activity as has been previously reported (Murakami et al., 2014). However, this M2 activity ramping did not differ based on whether that press would go on to exceed the criteria duration (Met) or not (Fail) (permutation testing that required 4 adjacent samples to pass the threshold for significance (Jean-Richard-dit-Bressel et al., 2020)). M2 activity during the lever press was modulated by whether that lever press would or would not meet the criteria (Figure 2.4C). This difference persisted following lever release, where there was an abrupt increase in Ca^{2+} activity just after the offset of Met presses, - i.e. reward delivery - as well as a subsequent sustained decrease in activity (Figure 2.4D). Thus M2 activity is modulated during lever pressing with ongoing modulation reflecting future success.

To determine if M2 activity related to ongoing and prior behavior, we created LME models to predict Ca^{2+} activity during epochs of the current lever press given both the ongoing action (press n duration) and prior behavior (the duration of press n - 1 to n - 6). We included prior activity as a covariate to control for autocorrelation in Ca^{2+} activity and compared β coefficients to 1000 order shuffled datasets. Before the onset of press n, there was a significant positive relationship between M2 activity and the just prior press durations (Figure 2.4E; n - 1, $p < 0.001$; n - 2, $p = 0.001$) and a small negative relationship between activity and the upcoming duration (press n, $p = 0.048$).

This representation of both current and prior lever press duration in M2 Ca^{2+} activity continued during the press itself (Figure 2.4F; press n, $p < 0.001$; n - 1: $p < 0.001$). At press offset, there was no relationship with the just completed press (n), but there was a positive relationship with the previous lever press duration (Figure 2.4G; n - 1, $p < 0.001$). We used complex LME models to investigate representation of other types of experiential information. M2 activity reflected Checking, IPI, and Time in Session, and these terms also interacted with the contribution of prior durations (Table 2.S5). In particular, at each lever press epoch we observed a significant, positive interaction between n - 1 duration and HE checking in between press n and n - 1, suggesting that checking increased the relationship between lever press duration and M2 activity. These complex LMEs were also better at predicting M2 Ca^{2+} activity relative to the simple LMEs that only included durations (difference in simple/complex prediction %: Before Press: +13.7%, During Press: +9.2%, After Press: +16.4%), showing that these often neglected variables are powerful drivers of M2 activity. Thus, we see representation of diverse aspects of subjective experience, aspects whose contributions are lost when M2 is lesioned. This suggests M2 circuits are recruited when a broad array of experiential information is used to guide behavior (as is often seen in exploration), but not when behavior can be accomplished using less flexible (i.e. habitual/model-free) processes.

Figure 2.4

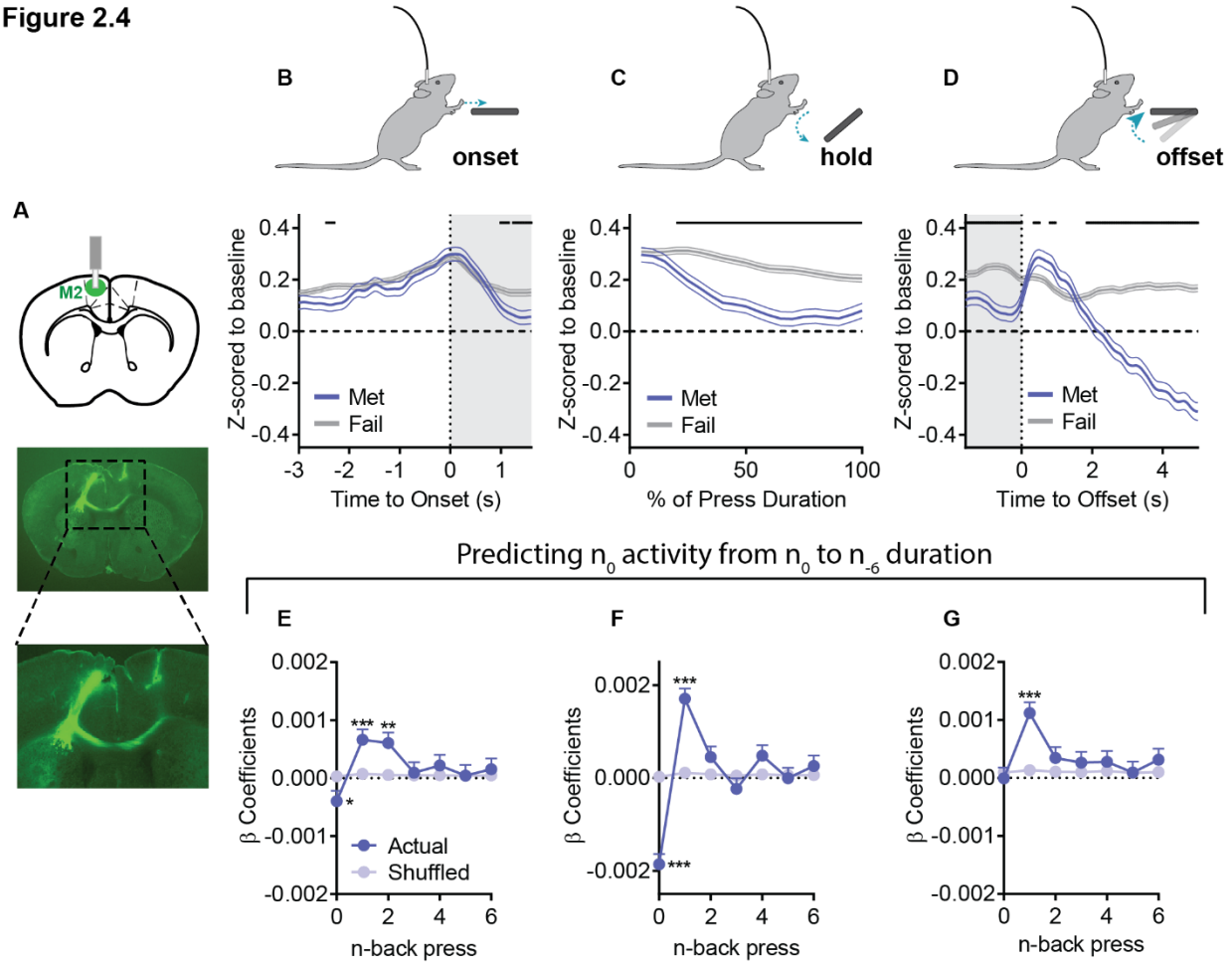


Figure 2.4. *M2* Ca^{2+} *activity represents prior experience*. **(A)** (top) Schematic and (bottom) example histology of M2 *in vivo* Ca^{2+} fiber photometry recordings. **(B-D)** Ca^{2+} activity z-scored relative to a baseline period and aligned to **(B)** press onset, **(C)** the hold down period itself (presented as the relative % of a press's duration), and **(D)** the offset of a lever press. **(E-G)** β coefficients from LME models relating activity to current and prior durations for Actual and order Shuffled data **(E)** before press onset, **(F)** during the press, and **(G)** after press offset. Met = Presses that met criterion, Fail = Presses that did not meet criterion. Grey shading in B, D indicates 1.6s. Black lines in B-D indicate significant differences between Met/Fail via permutation testing. Shuffled data are the mean+SEM of 1000 order shuffled β coefficients. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. See also Table 2.S5.

M2-DMS projections use recent experience to plan upcoming actions

M2 sends dense projections into dorsal medial striatum (M2-DMS) regions (Delevich et al., 2020; Hintiryan et al., 2016) that contribute to action selection (Klaus et al., 2019), but it is unclear what information is conveyed. We performed *in vivo* fiber photometry of virally targeted M2-DMS activity and examined representation of experiential information within this population (n = 7, Figure 2.5A). We again observed a ramping in M2-DMS Ca^{2+} activity prior to lever press onset. This activity reflected future success, with larger increases in activity for presses that *would* meet criteria (Figure 2.5B). This relationship was also present during the press itself (Figure 2.5C), and upon lever release (Figure 2.5D), raising the hypothesis that M2-DMS projections may carry information specifying and/or planning actions based on prior experience.

Indeed, LMEs using durations to predict M2-DMS activity before press onset showed both prior (Figure 2.5E; n - 1, $p < 0.001$) and upcoming (n, $p < 0.001$) durations were positively related to M2-DMS Ca^{2+} activity. M2-DMS activity during the press did not relate to the current duration, but was positively related to the prior duration (Figure 2.5F; n - 1, $p < 0.001$), and likewise at press offset there was a positive relationship only with the n - 1 press duration (Figure 2.5G; $p < 0.001$). Furthermore, complex LMEs revealed significant influences of Checking, prior Reward, IPI, and Time in Session on M2-DMS activity (Table 2.S6). As in M2, at every time point there was an interaction between checking and n - 1 duration on M2-DMS activity. Further, the complex LMEs predicted more of the data relative to the simple models (difference in simple/complex prediction %: Before Press: +13.2%, During

Press: +9.2%, After Press: +15.2%). Thus M2-DMS activity appears to reflect the use of recent experiences to plan upcoming actions.

Figure 2.5

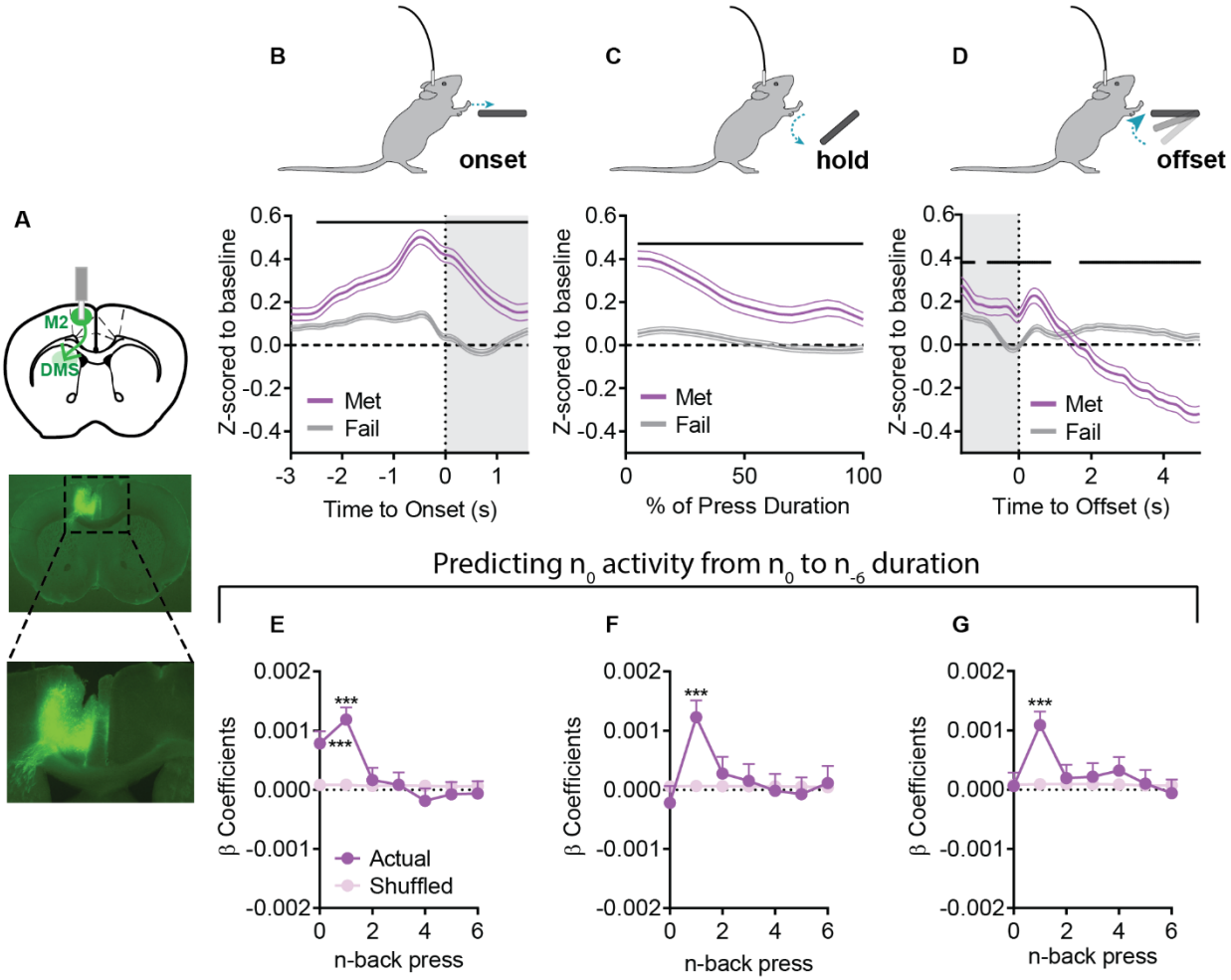
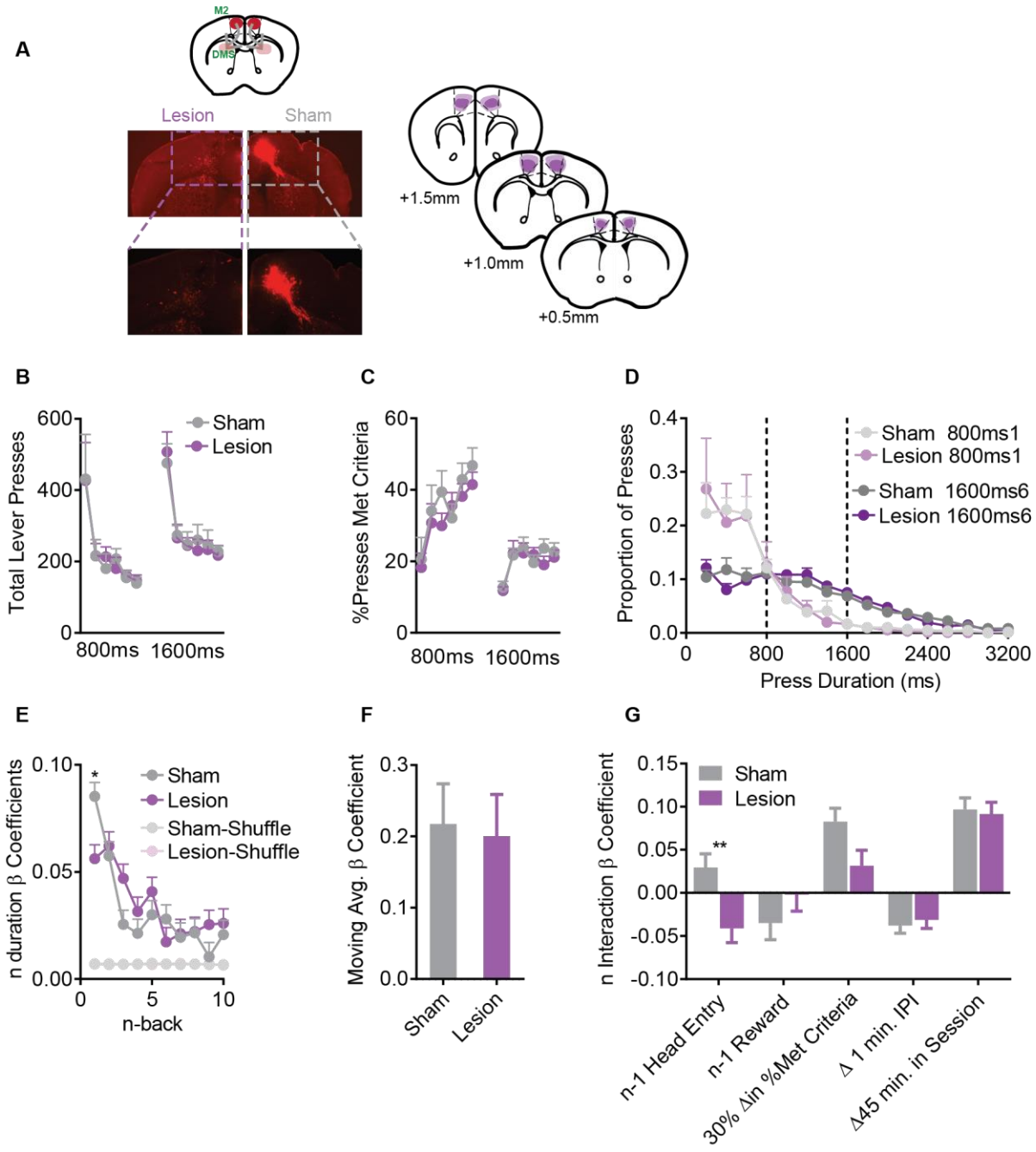


Figure 2.5. *M2-DMS* Ca^{2+} *activity encodes preceding and upcoming actions*. **(A)** (top) Schematic and (bottom) example histology of projection specific M2-DMS Ca^{2+} fiber photometry. **(B-D)** Ca^{2+} activity z-scored relative to baseline and aligned to **(B)** press onset, **(C)** the duration of the press, and **(D)** press offset. **(E-G)** β coefficients from LME models relating activity to current and prior durations for Actual and Shuffled data **(E)** before press onset, **(F)** during the press, and **(G)** after press offset. Met = Presses that met criteria. Fail = Presses that did not meet criterion. Grey shading in B, D indicates 1.6s. Black lines in B-D indicate significant differences between Met/Fail via permutation testing. Shuffled data are the mean+SEM of 1000 order shuffled β coefficients. *** $p < 0.001$. See also Table 2.S6.

To test whether M2-DMS activity functionally contributed to planning actions based on recent experience, we used a Cre-dependent caspase strategy to selectively lesion M2-DMS projection neurons prior to training (Figure 2.6A, n = 8 Lesion, n = 8 Sham). Again, we observed no effect on coarse measures of behavior including Total Lever Presses, %Presses Met Criteria, and Press Durations (Figures 2.6B-D). Simple LME modeling showed M2-DMS lesions reduced the relationship between press n and press n - 1 (Figure 2.6E; 2-way ANOVA (n-back x Sham/Lesion) main effect of n-back ($F_{9,467760}$= 14.6, $p < 0.0001$); significant interaction ($F_{9,467760}$= 2.29, $p = 0.0144$)). This deficit was selective to n - 1 (multiple comparison corrected post-hoc n - 1: $t_{467760} = 3.09$, $p = 0.021$). There was no effect on the Moving Average (Figure 2.6F). Interestingly, complex LMEs revealed a more specific deficit in M2-DMS lesions relative to broad M2 populations; M2-DMS lesions reversed the contribution of a checking HE between press n and press n - 1, such that checking was now detrimental in the use of prior duration information to guide performance (Figure 2.6G; $t_{47352} = 3.10$, $p = 0.002$). However, no other terms differed between Sham and Lesion groups (Table 2.S7).

Figure 2.6. *Pretraining M2-DMS lesions impair use of recent experience.* **(A)** (top) Schematic, (bottom) example histology and (right) average/maximal (dark/light shading) spread of projection specific M2-DMS lesion using a Cre-dependent caspase strategy, slice coordinates relative to Bregma. **(B)** Total Lever Presses across training days. **(C)** %Presses that met criteria across training days. **(D)** Histogram of lever press durations on the first and last days of the 800ms and 1600ms criteria (200ms bins). **(E)** β coefficient from LME models predicting n duration from n-back durations for Actual and order Shuffled data. **(F)** Moving average β coefficient for Actual and Shuffled data. **(G)** β coefficients for the interaction terms from the complex LME model. For display purposes we transformed the continuous variables to show relevant changes. B-D are mean+SEM across mice. Shuffled data are the mean+SEM of 1000 order shuffled β coefficients. ** $p < 0.01$, * $p < 0.05$. See also Table 2.S7.

Figure 2.6


A
M2
DMS
Lesion
Sham
+1.5mm
+1.0mm
+0.5mm
B
Total Lever Presses
600
400
200
0
800ms 1600ms
Sham
Lesion
C
%Presses Met Criteria
60
40
20
0
800ms 1600ms
D
Proportion of Presses
0.4
0.3
0.2
0.1
0.0
0
800
1600
2400
3200
Press Duration (ms)
Sham 800ms1
Lesion 800ms1
Sham 1600ms6
Lesion 1600ms6
E
n duration β Coefficients
0.10
0.05
0.00
0
5
10
n-back
*
Sham
Lesion
Sham-Shuffle
Lesion-Shuffle
F
Moving Avg. β Coefficient
0.3
0.2
0.1
0.0
Sham
Lesion
G
n Interaction β Coefficient
0.15
0.10
0.05
0.00
-0.05
-0.10
**
Sham
Lesion
n-1 Head Entry
n-1 Reward
30% Δin %Met Criteria
Δ 1 min. IPI
Δ45 min. in Session

The photometry and lesion data together suggest that M2-DMS activity represents and is functionally necessary for recent sequential action (pressing and checking) experience to contribute to the initiation and execution of the current decision. To directly test this hypothesis, we took a behaviorally-dependent, closed-loop optogenetic approach to inhibit M2-DMS neural activity. We used a dual-virus strategy to express an inhibitory opsin (ArchT: n = 5, Figure 2.7A) that reduced M2-DMS spiking when activated by light (Figure 2.7B). We targeted inhibition to three different epochs: the initiation of a lever press, during the press itself, and after lever press release. Each manipulation occurred across 6 days of training and only on a subset of lever presses. This allowed us to include additional terms in our LME models to determine if inhibition directly affected press n duration, and/or if inhibition affected the contribution of prior experience (i.e., an interaction between inhibition and n - 1 duration). In addition to this within subject comparison, we also made between subject comparisons to fluorophore control mice (tdTomato: n = 6).

In order to target inhibition prior to press onset, mice were tracked using an overhead camera and light (1s, continuous) was triggered 50% of the time when mice entered a zone centered on the lever. We did not find any effect of pre-onset M2-DMS inhibition on overall performance (Figure 2.7D), nor any effect on press duration itself (i.e., no main effect of inhibition, Figure 2.7E). Within the ArchT group, inhibition prior to lever pressing did induce a significant negative interaction with n - 1 duration ($F_{1,2718}$ = 10.6, p = 0.001), and a significant difference with the tdTomato group (Figure 2.7F; t_{6645} = 2.04, p = 0.042). However, as this inhibition continued for 1s, it may have persisted during lever pressing itself. Indeed, there was no longer an effect of inhibition

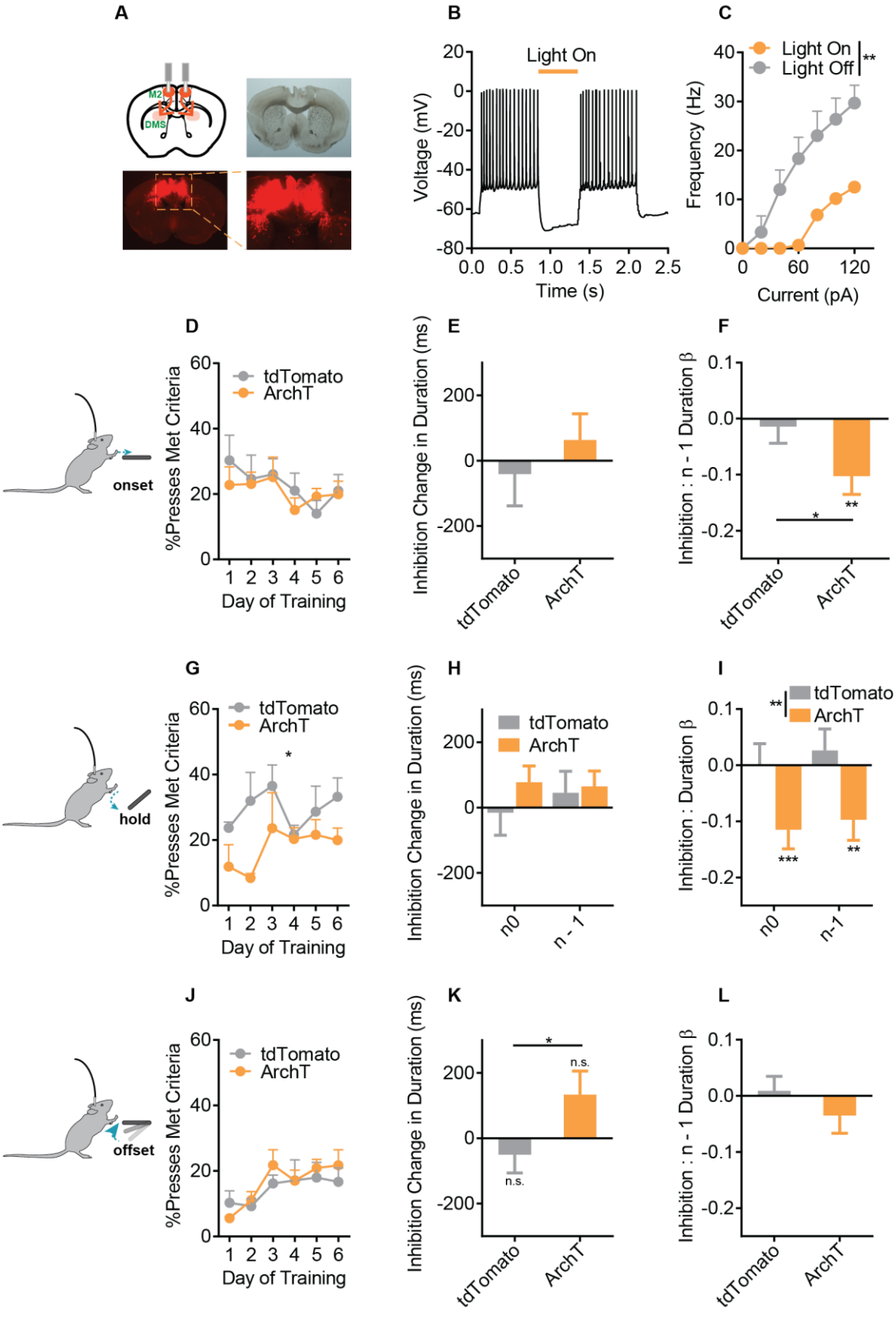
even within the ArchT group ($F_{1,534}$ = 0.45, p = 0.50) when we limited our analysis to light inhibition that did not spillover into the lever press. Thus, it seems likely that the marginal effect of pre-onset inhibition is due to spillover of inhibition to the lever press itself as opposed to inhibition that occurred prior to press onset.

To directly address this question, we next tied inhibition to lever pressing itself. We inhibited during the full duration of every 7th lever press. Such inhibition did reduce the efficacy of lever pressing (Figure 2.7G; 2-way RM-ANOVA (Opsin/Fluorophore X Day), main effect only of Opsin group ($F_{1,9}$ = 7.59, p = 0.0223)). Again there was no main effect of inhibition on press duration itself, and this was true both whether inhibition occurred during press n or on press n - 1 (Figure 2.7H). However, M2-DMS inhibition during press n prevented the use of n - 1 duration information from guiding the current action (Figure 2.7I). Further, inhibition during press n also prevented the experiential information gained during the execution of press n from informing the next press (F-test within ArchT model, n0: $F_{1,6110}$ = 11.2, p = 0.0008; n - 1: $F_{1,6110}$ = 6.91, p = 0.009. Group comparison between ArchT/tdTomato 2-way ANOVA (Opsin x n-back) main effect only of Opsin group $F_{1,23602}$ = 10.5, p = 0.0012). This suggests that M2-DMS activity during the press itself is not important for controlling the duration of the current lever press *per se*. Instead, this activity contributes to using recent experiential information to execute current actions, and this abrupt disruption impaired task performance. In support of this, when we targeted inhibition to press offset (1s of light after press release), there was neither a direct effect of inhibition on subsequent press durations, nor an interaction with the use of recent experience, nor any effect on performance (Figures 2.7J-L). There was also no interaction with the moving average

term at any inhibition time point. The lack of a direct effect of inhibition on press duration suggests the deficit in use of recent experience was not due to a non-specific motor effect. The lack of any inhibition effect prior to or after execution of the lever press also suggests M2-DMS activity does not represent a form of working memory, but instead supports use of prior experience to inform action execution.

Figure 2.7. *Optogenetic inhibition of M2-DMS projections during execution impairs use of recent experience.* **(A)** (top left) Schematic and example histology of ArchT optogenetic inhibition of M2-DMS projection neurons. **(B-C)** Slice verification of ArchT-mediated inhibition in M2-DMS projection neurons. ** = 2-way RM ANOVA (Current x Light) interaction: $F_{6,6} = 17.0$, $p = 0.002$. **(D-F)** Pre-onset inhibition. **(D)** %Presses that met criteria, **(E)** Main effect of inhibition on duration, and **(F)** Interaction between inhibition and the contribution of the prior duration. **(G-I)** As in D-F except for inhibition during the duration of the press. **(J-L)** As in D-F except for inhibition occurring after press offset. Note that although there is a group difference between ArchT and tdTomato mice in K, the main effect itself is n.s. in the model for both groups. n0 = Light occurred on press n. n - 1 = Light occurred on press n - 1. tdTomato = tdTomato expressing control mice. ArchT = ArchT expressing mice. In all LME graphs, * without any lines indicate significant terms via F-test on the model (within group comparison), while * with a line indicate significant, between group differences. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. n.s. = Not significant. Data are mean+SEM.

Figure 2.7


A
M2
DMS
B
Light On
Voltage (mV)
Time (s)
C
Light On
Light Off
Frequency (Hz)
Current (pA)
D
onset
%Presses Met Criteria
tdTomato
ArchT
Day of Training
E
Inhibition Change in Duration (ms)
F
Inhibition : n - 1 Duration β
G
hold
H
I
Inhibition : Duration β
J
offset
K
n.s.
L

Discussion

There is a growing concern that neuroscience investigations into decision-making are “missing the forest for the trees”, or *vice versa*. Investigations into the nature of decision-making that isolate specific task-based computations or focus on summary statistics such as accuracy have been indispensable in providing information about both the tree and forest, respectively. However, the present data suggests the need to account for the mesoscopic context experiential information provides in order to link these levels of analysis, akin to understanding the intertwined communication among trees in a forest (Gorzelak et al., 2015). Here, sources of experiential information often treated as task-irrelevant determined whether and to what degree recent experience-information influenced adaptive behavior to support ongoing decision-making. Experiential information influenced the neural correlates of decision-making and determined circuit recruitment and contribution in mice, suggesting such information should not be ignored. By using this approach, we show that M2 and M2-DMS circuits use broad experiential information to instigate exploratory or recent experience-based responding.

Classic temporal difference or reinforcement learning models (Rescorla & Wagner, 1972; Sutton & Barto, 1998) emphasize the role played by responses and outcomes. By using an unconstrained task with a continuous decision variable, we found mice do not drastically shift their strategy solely on their sequence of actions, nor based on whether their action earned a reward. Rather, experiential variables such as checking and the passage of time strongly influenced the behavior and may serve to arbitrate between strategies. Behaviorally the latter could function to bias

exploratory responding when the preceding action is more distant in time and hence, when the environment, and/or its neural representation may have changed. On the other hand, the former suggests information seeking itself increases use of experience-based strategies, perhaps as a result of providing definitive feedback. As multiple behavioral controllers can be used to make seemingly similar decisions (Bouton & Balleine, 2019), experiential information may be used to bias strategy-level recruitment for instance, by adjusting the relative degree of exploration (Figure 2.3) or the relative similarity between adjacent decisions (Figures 2.6-7). This bias in recruitment strategy may arise through experiential modulation of associated neural activity (Schreiner & Gremel, 2018), perhaps by setting the "gain" on behavioral strategies (Johnson et al., 2016).

Integration and implementation of subjective experience in M2 and M2-DMS circuits

We found a robust representation of subjective experience in broader M2 population activity. That lesions both increase the similarity between adjacent presses and reduce the integration of non-action-outcome information (including Checking, IPI, Time, and overall performance) suggests M2 lesions may render mice relatively inflexible and left to rely more on a simple repetition-based strategy. Though it must be noted that the M2 Sham comparison group displays a low n - 1 β coefficient, perhaps inflating the overall effect of M2 Lesions. However, the increased relationship is still present when comparisons are made to another sham group (M2-DMS Shams; t_{56426} = 2.22, p = 0.026), with a similar loss of interactions with other experiential information. Collectively, our results extend previous studies implicating M2 populations in goal-

directed or model-based decision-making (Gremel & Costa, 2013) by providing novel insight into precisely *how* this effect is achieved. Namely, by nullifying the contribution of subjective experience in arbitrating between decision-strategies, animals with M2 lesions rely on repetition-based strategies.

While M2 is important for a broader representation of experiential information, in a subset of M2 projections neurons (M2-DMS) we see a more limited contribution of information used to guide ongoing actions. Converging Ca^{2+} activity, lesion, and optogenetic inhibition studies implicate M2-DMS projections specifically in contributing recent action information to ongoing actions. The reduction in the use of recent experience only when optogenetic inhibition occurred *during* the press suggests that M2-DMS activity may serve as an experience-based guide or reference for ongoing actions. This raises the hypothesis that M2-DMS may function as a comparator for template or pattern matching during action performance, analogous to the pattern matching seen in avian vocal learning, and hypothesized to be implemented in premotor regions (Mooney, 2009). In M2-DMS lesioned mice, an intermediate behavior of checking in between lever presses reduced the reliance of the current action on the prior, providing some evidence that M2-DMS function is necessary to link and/or compare recent action experience as has been suggested by work examining sequence learning and initiation (Rothwell et al., 2015). Future studies investigating M2-DMS function at the single neuron level could reveal important insights into precisely how this is instantiated in the brain, and if there is an “embodied engram” of recent actions, or a comparator function in M2-DMS projection neurons.

Conclusion

Rarely is behavior in the natural world so neatly constrained as in many laboratory tasks; thus it seems likely that animals have adapted to use diverse sources of information to guide their behavior. The brain should therefore be sensitive to this information, yet several recent studies have demonstrated remarkably widespread coding of variables in the brain (Allen et al., 2017; Steinmetz et al., 2019). Perhaps this apparent distributed coding is the consequence of attributing relatively static measurements of behavior and human-derived constructs to large neural populations. That there is a wealth of information available to animals and many neural circuits to support decision-making, raises the hypothesis that specific aspects of experiential information may modulate neural function differentially depending on the circuit and the computation. Indeed, we find that M2 is sensitive to many aspects of this experiential information, but examination of a discrete output population (M2-DMS) showed a more selective representation and functional role. Investigations into circuit, synaptic, and molecular mechanisms controlling how subjective experience modulates decision-making will likely be fruitful, akin to increased understanding of arousal modulation of sensory processing (Shimaoka et al., 2018).

Repetitive decision-making is found across many disease states including substance use disorders and Obsessive Compulsive Disorder. M2's potential human homologues - the pre-supplementary/supplementary motor areas - are accessible to region-specific treatments such as TMS, which have shown promise in disease treatment (Gomes et al., 2012; Hawken et al., 2016; Mantovani et al., 2013). Here we

establish that M2-DMS is involved in implementing repetitive or recent-experience-based decisions. This raises the hypothesis that M2-DMS dysfunction may lead to decisions that are inappropriately or excessively repetitive (Corbit et al., 2019). Incorporating subjective experience into the examination of disease-induced brain function during decision-making may increase the likelihood of obtaining enduring findings relevant to the clinical treatment of disease.

Methods

Experimental Model and Subject Details

Similar numbers of male and female C57BL/6J mice (>7 weeks/50 PND) (The Jackson Laboratory, Bar Harbour, ME) were used for experiments. Exploratory analyses for sex differences in the behavioral cohort revealed no differences, and thus we collapsed across sex. All procedures were conducted during the light period and mice had free access to water throughout the experiment. Mice were housed 2–4 per cage on a 14:10 light:dark cycle. Mice were at least 6 weeks of age prior to surgery. Mice were food restricted to 85-90% of their baseline weight 2 days prior to the start of behavioral procedures, and were fed 1–4 hours after the daily training. All experiments were approved by the University of California San Diego Institutional Animal Care and Use Committee and were carried out in accordance with the National Institutes of Health (NIH) “Principles of Laboratory Care”.

Behavioral Procedures

Mice were trained once per day in operant chambers in sound attenuating boxes (Med-Associates, St Albans, VT) in which they pressed a lever (left or right of the food magazine, counterbalanced for location) for an outcome of regular ‘chow’ pellets (20 mg pellet per reinforcer, Bio-Serv formula F0071). Each training session commenced with an illumination of the house light and lever extension and ended after either 60 reinforcers were earned or 90 minutes had elapsed, with the lever retracting and the house light turning off.

On the first day, mice were trained to approach the food magazine to retrieve the pellet outcome (no lever present) on a random time (RT) schedule, with a reinforcer delivered on average every 120 seconds for a total of 60 minutes. Next, mice were trained on a continuous ratio schedule of reinforcement (CRF) across 3 days, where every lever press was reinforced (no duration requirement), with the total possible number of reinforcers increasing (CRF10, 30, and 60).

Following CRF pretraining, the hold down task was introduced. We instituted a duration requirement on lever pressing: animals had to press and hold down the lever for >800ms in order to earn food reward (delivered immediately after press release). Importantly, there were no cues, no timeout period, nor any discrete trials; the lever was always available to mice, until they completed their session. Mice were trained at the >800ms criterion for 6 days, followed by at least 6 days of training at the >1600ms criterion. During all days, timestamps for lever press onset, lever press offset, the onset and offset of headentry into the food magazine, and the delivery of reward were recorded. From this timing information, we were able to calculate durations of lever

presses and headentries. Of note, use of Med Associates introduced a 20 ms limit on our time resolution.

Outcome Devaluation

In the behavioral mice (Figure 2.1 and Figure 2.S1, n = 12 total, n = 7 female and n = 5 male), after 8 days of training at >1600ms, we performed outcome specific satiety. Devaluation procedures occurred across two days. In brief, on the valued day, mice had *ad libitum* access to an outcome previously experienced in the home cage for 1 hour before being placed in the training context for a 5 minute, non-reinforced test session. On the devalued day, mice were given 1 hour of *ad libitum* access to the outcome previously earned by lever press, and then underwent a 5 minute, non-reinforced test session in the training context. One mouse consumed less than 0.1g of the valued outcome during pretraining exposure and was excluded from all devaluation analyses (giving final n = 11). The order of revaluation day was counterbalanced across mice.

Probabilistic Reward

A naive group of mice (n = 15, n = 4 female and n = 11 male) were trained for 6 days on >800 ms, followed by 8 days at > 1600ms criteria, and then switched to probabilistic reward, where only a percentage of presses that exceeded the duration criterion were rewarded on a random ratio schedule. These animals were separated into three different probabilistic reward groups: 25%, 50%, and 75% (n = 5 each group) and trained for a further 3 days under the assigned probabilistic schedule.

Surgical Procedures

All viral vectors were obtained from the UNC Viral Vector Core (Chapel Hill, NC) or Addgene (Wateron, MA). Mice were anaesthetized with isoflurane (1–2%) and intracranial injections were performed via Hamilton syringe (Reno, NV) targeted at a relatively posterior portion of M2 (from Bregma: AP +1.0mm, L ±0.5mm and V −1.2mm, -1.4mm from the skull), and/or DMS (from Bregma: AP +1.0mm, L ±1.65mm and V −3.0mm, -3.2mm from the skull). Syringes were left in place for five minutes after each injection to allow for diffusion, and all viruses or drugs were infused at a rate of 100nl/min. Mice were given at least two weeks to allow for recovery and viral expression before the start of experimental procedures (at least four weeks for all M2-DMS manipulations). After behavioral testing was concluded, mice were euthanized and brains were extracted and fixed in 4% paraformaldehyde. Localization and spread of viral expression was assessed in 50-100 μm thick brain slices using fluorescent microscopy (Olympus MVX10, Shinjuku, Japan).

For M2 lesions, n = 12 Lesion mice were bilaterally injected with ibotenic acid (10mg/ml, ThermoFisher), while n = 12 Sham lesion mice were injected with vehicle (saline) at M2 (2 injections of 120nl at V -1.4mm and -1.2mm from the skull in each hemisphere). In order to assess excitotoxic lesion presence and spread, brains were sliced at 50um thick, and incubated with propidium iodide (1:10000 in 1xPBS, Chemodex: P0023) and Isolectin-GS IB_4 Alexa Fluor 488 Conjugate (20:10000, ThermoFisher: I21411), a marker of microglial cells which are recruited via lesions (Lünemann et al., 2006). Brain slices were incubated for 1hr, followed by 3 x 10min

washes. 4 Sham mice were excluded due to technical difficulties during training, and 2 Lesion mice were excluded due to histology, giving final n's of 10 Lesion (n = 6 female, n = 4 male) and 8 Sham (n = 4 female, n = 4 male) mice.

For M2 GCaMP experiments, n = 8 mice (n = 4 female, n = 4 male) were injected (2 injections of 200nl at V -1.4mm and 1.2mm from the skull) with rAAVDJ/PAAV-CaMKIIa-GCaMP6s to express GCaMP6s under control of the Ca^{2+} calmodulin dependent protein kinase IIα (CamKIIα) promoter and implanted with an optical fiber unilaterally in M2.

For M2-DMS GCaMP experiments, n = 8 mice were unilaterally injected with a viral vector expressing Cre recombinase (AAV5/Ef1a-Cre-WPRE) in DMS (2 injection depths: V -3.0mm and -2.8mm from the skull, 250nl each), and were injected with a viral vector expressing a Cre-dependent GCaMP6s (pAAV.CAG.FLEX.GCaMP6s.WPRE.SV40 (Addgene: 100842); 2 injection depths: V: -1.4mm, and -1.2mm from the skull, 200nl each) followed by fiber implantation in ipsilateral M2. One mouse was excluded due to histology (n = 4 female, n = 3 male).

For M2-DMS lesion, n = 8 Lesion (n = 4 female, n = 4 male) and n = 8 Sham (n = 4 female, n = 4 male) mice were bilaterally injected with 200nl of a viral vector expressing CamKIIα-Cre in DMS (rAAV5/CamKII-GFP-Cre; 2 injection depths: V: -3.1mm and -2.9mm from skull, 200nl each). Lesion and Sham mice were also injected with a viral vector expressing Cre-dependent tdTomato in M2 (rAAV5/Flex-tdTomato; 100nl at V -1.4mm from the skull). Lesion mice additionally received a viral vector expressing a Cre-dependent caspase virus in M2 to induce apoptosis of M2-DMS

projections (rAAV5/AAV-Flex-taCasP3-TEVP; 2 injection depths: V -1.4mm and -1.2mm from the skull, 200nl each).

For M2-DMS optogenetic inhibition experiments, n = 8 Archaerhodopsin (ArchT) and n = 8 tdTomato mice were bilaterally injected with a viral vector expressing CamKIIα-Cre in DMS (rAAV5/CamKII-GFP-Cre; 2 injection depths: V -3.1mm and -2.9mm from the skull, 200nl each). Following exclusion for viral expression or low levels of behavior, there were n = 5 ArchT mice (n = 3 male, n = 2 female), and n = 6 tdTomato control mice (n = 3 male, n = 3 female). Due to the proximity of bilateral M2 at this posterior portion (~1.0mm) for ferrule implantation, we injected virus and implanted fibers at a 10° angle, and adjusted the M2 coordinates accordingly. Experimental ArchT mice received a viral vector expressing a Cre-dependent inhibitory opsin (rAAV5/Flex-ArchT-tdTomato), while fluorophore control mice received a viral vector expressing Cre-dependent fluorophore only (rAAV5/Flex-tdTomato), in both cases receiving the same injection volume (300nl at V -1.42mm from the skull), with bilateral M2 fibers implanted at V -1.37mm from the skull.

Fiber Photometry

Animals underwent pre-training as described above, but received one additional day of CRF training during which animals were first hooked up to 400 um optical fiber tethers with ferrule to ferrule connectivity. A 470nm LED (Thorlabs, Newton, NJ) was used for excitation of GCaMP6s (< 70 µW/mm2), and fluorescence emissions were collected with a bifurcated fiber (Thorlabs, Newton, NJ) which allowed for simultaneous, independent recordings of two mice. We imaged the dual fiber core

using a 4x objective (Olympus, Shinjuku, Japan) focused onto a CMOS camera (FLIR Systems, Wilsonville, OR). Regions of interest (i.e., the fiber cores) were selected using Bonsai software (Lopes et al., 2015) to acquire fluorescence intensity signals (at a rate of 20Hz). Bonsai software simultaneously collected analog behavioral data and timestamps for lever presses, head entries, and reinforcer delivery sent via TTL Med-PC pulses using microprocessors (Arduino Duo, from Arduino, Sumerville, MA) with custom code. Photometry and behavioral data were imported into Matlab (Mathworks Inc., Natick, MA) for analysis using custom scripts. To account for decay across the session (photobleaching), we fit the fluorescence intensity signal to a double exponential. To check for bad coupling of the fiber to the ferrule, or low expression each session we calculated the 97.5 percentile of dF/F and ensured that there was at least a 1% change, sessions failing to meet this criterion were excluded from analyses (Markowitz et al., 2018), and also excluded sessions with visual anomalies in the session long traces (e.g., a sudden, sustained decrease in activity partway through the session that could indicate fiber decoupling). We used the mean and standard deviation during a baseline period -15s to -5s prior to lever pressing to z-score press-related activity (i.e., from -5s prior to onset up to 5s post offset). To compare Met and Fail lever presses, we performed running permutation tests, requiring that at least 4 adjacent samples were significantly different from one another to control for fluctuations in the data (functions implemented in Matlab from (Jean-Richard-dit-Bressel et al., 2020)). We smoothed Ca^{2+} activity data using a 10 sample (or 5 sample for interpolated activity) long Gaussian filter for display purposes only.

Optogenetic Inhibition

For bilateral light delivery, Arduino Duos with custom code were used to receive TTL pulses from Med-PC operant chambers and trigger onset of 2 LEDs (595nm, Thorlabs) coupled to 200um sheathed fiber optic cable with ferrule to ferrule connectivity (>= 1mW at ferrule tip). We used 595nm light as this has been shown to optimally activate ArchT while avoiding non-specific effects (Setsuie et al., 2020). We used several different protocols to target the closed-loop inhibition to different task epochs. Inhibition during the duration of the lever press occurred across the 6 >800ms training days, with light delivery (continuous, not pulsed) tied to the lever pressing itself. As we observed a decaying relationship between n-back press durations and n - 0 press duration, every 7th lever press triggered light delivery, which persisted for the duration of the lever press (with a time resolution of 20ms for light offset). During days 1-6 of the >1600ms training, we instead tied light delivery to press *offset*, again, on every 7th lever press. Thus, after every 7th lever press, mice were given 1s of light. Finally, after undergoing 4 days of baseline training without any light inhibition (though while still being hooked up to fibers), we shifted to inhibiting *prior* to press onset for 6 days. In order to achieve this, animals were recorded with an overhead camera (1080p wide angle webcam, Logitech) and tracked in real time using Bonsai software. We individually defined regions of interest centered on the lever (approximately twice the width and length of the lever itself). 50% of entrances into this region generated a TTL pulse to turn on the LEDs, which remained on for 1s.

ArchT Slice Validation

Coronal slices (250 μm thick) containing M2 were prepared using a Pelco easiSlicer (Ted Pella Inc, Redding, CA). Mice were anesthetized by inhalation of isoflurane and brains were rapidly removed and placed in 4°C oxygenated ACSF containing the following (in mM): 210 sucrose, 26.2 $NaHCO_3$, 1 NaH_2PO_4, 2.5 KCl, 11 dextrose, bubbled with 95% O_2/5% CO_2. Slices were transferred to an ACSF solution for incubation containing the following (in mM): 120 NaCl, 25 $NaHCO_3$, 1.23 NaH_2PO_4, 3.3 KCl, 2.4 $MgCl_2$, 1.8 $CaCl_2$, 10 dextrose. Slices were continuously bubbled with 95% O_2/5% CO_2 at pH 7.4, 32°C and were maintained in this solution for at least 60 min prior to recording.

Whole-cell current clamp recordings were made in pyramidal cells of M2. Pyramidal cells that expressed ArchT were identified by the fluorescent tdTomato label using an Olympus BX51WI microscope mounted on a vibration isolation table and a high-power LED (LED4D067, Thorlabs). Recordings were made in ACSF containing (in mM): 120 NaCl, 25 $NaHCO_3$, 1.23 NaH_2PO_4, 3.3 KCl, 0.9 $MgCl_2$, 2.0 $CaCl_2$, and 10 dextrose, bubbled with 95% O_2/5% CO_2. ACSF was continuously perfused at a rate of 2.0 mL/min and maintained at a temperature of 32°C. Picrotoxin (50 μM) was included in the recording ACSF to block GABAA receptor-mediated synaptic currents. Recording electrodes (thin-wall glass, WPI Instruments) were made using a PC-10 puller (Narishige International, Amityville, NY) to yield resistances between 3–6 MΩ. Electrodes were filled with (in mM): 135 $KMeSO_4$, 12 NaCl, 0.5 EGTA, 10 HEPES, 2 Mg-ATP, 0.3 Tris-GTP, 260–270 mOsm (pH 7.3). Access resistance was monitored throughout the experiments. Cells in which access resistance varied more than 20% were not included in the analysis.

Recordings were made using a MultiClamp 700B amplifier (Molecular Devices, Union City, CA), filtered at 2 kHz, digitized at 10 kHz with Instrutech ITC-18 (HEKA Instruments, Bellmore, NY), and displayed and saved using AxographX (Axograph, Sydney, Australia). A series of fixed current injections (20 pA increments from 0 to 120 pA) were used to elicit action potential firing and the number of spikes were counted at each current step. For verification of ArchT function, ArchT was optically stimulated using 590nm light, delivered via field illumination using a high-power LED (LED4D067, Thorlabs). Optical stimulation was done under constant illumination for 1s during current injections.

Data Analysis

Linear Mixed Effects Models of Behavior

We built simple Linear Mixed Effects (LME) models to model the relationship between the duration of lever press n and n-back (n - 1 through n - 10) lever press durations. We included random intercept terms for mouse and day of training to account for the repeated structure of our data. To determine how far back a significant relationship existed between press n and any particular n-back press, we shuffled the order of a particular n-back (e.g., only n - 3) 1000 times and compared the shuffled distribution of beta coefficients to the actual value via permutation test. Of note, we are shuffling here within individual mouse/sessions, thus preserving the overall statistics of the data, and shuffling only the order in which a specific type of event occurred.

$$n = \beta_0 + \beta_{n-1}n_{-1} + \beta_{n-2}n_{-2} + \ldots + \beta_{n-10}n_{-10} + \beta_t(t) + \beta_{\%}(\%) + (1|M) + (1|D) + \varepsilon_i$$

Where n is the current lever press duration, n - 1 through n - 10 are the previous 1 through 10 lever press durations and β_x is the linear regression coefficient for term x (β_0 being the intercept term). We also included covariates of time in session (t) and the percentage of presses that met criteria (%). We included random intercept terms for both mouse (M) and day (D).

In order to determine which other experiential variables affect lever press n duration, we also built more complex LME models that included additional variables. To select variables for this model, we created a “full” model that included n-back durations up to n - 6 (as that is as far back as we see a consistent difference from shuffled data in the simple models), and then main effects of other variables and their interactions with n-back durations, also up to n - 6 (e.g., a binary for if mice made a checking headentry after the previous lever press). We individually removed terms from this full model, and compared Bayesian Information Criterion (BIC) to assess if adding a term improved the model. If any term did not improve the model, we removed it, and also removed any further n-back examples of it. However, we kept main effect terms in the model if the interactions were significant, and kept all the same interaction terms for n - 1 and the moving average term to be able to directly compare how various events might differentially affect the contribution of press n - 1 versus the moving average. To ensure that terms in this reduced model did not improve the model due to overall correlations across days or mice, we also compared beta coefficients from the actual data to those obtained from 1000 order shuffled datasets, where we individually permuted a given term within individual mouse/sessions. This analysis conducted on our “reduced” model agreed with the BIC selection for terms

that improved the model. We were ultimately left with the model in Table 2.S3 (see also Table 2.S2 for a description of the terms), signified by the equation below.

$$n = \beta_0 + \beta_{n-1}n_{-1} + \beta_{n-2}n_{-2} + \ldots + \beta_{n-6}n_{-6} + \beta_{MA}MA + \beta_t t + \beta_{\%}\% + \beta_{IPI}IPI_{-1} + \beta_{IPI}IPI_{-2} + \beta_R R_{-1} + \beta_{HE}HE_{-1} + \beta_{t*n-1}t*n\text{-}1 + \beta_{\%*n-1}\%*n\text{-}1 + \beta_{IPI*n-1}IPI_{-1}*n\text{-}1 + \beta_{IPI*n-2}IPI_{-2}*n\text{-}2 + \beta_{R*n-1}R_{-1}*n\text{-}1 + \beta_{HE*n-1}HE_{-1}*n\text{-}1 + \beta_{t*MA}t*MA + \beta_{\%*MA}\%*MA + \beta_{IPI*MA}IPI_{-1}*MA + \beta_{R*MA}R_{-1}MA + \beta_{HE*MA}HE_{-1}*MA + (1|M) + (1|D) + \varepsilon_i$$

Where β_x represents the linear regression coefficient for a given term. This model has the same terms as the simple model, though only back to n - 6 durations, as that is as far back as there is a reliable difference to shuffled data. In addition, there is the MA term which is a moving average from presses n - 7 through n - 60 (length selected via BIC using different window lengths). Additionally, we have main effects of time in session (t, in ms), the percentage of presses that met criteria (%), inter-press interval (IPI in ms, for both time in between press n and press n - 1 (IPI_{-1}), and between press n and press n - 2 (IPI_{-2})), outcome of press n - 1 (R_{-1}: binary where 0 is no reward and 1 is reward), and headentry between press n - 1 and press n (HE_{-1}: binary where 0 is no headentry and 1 is headentry). Again, we have random intercept terms for mouse (M) and day of training (D). We also included interaction terms between the n - 1 duration term and: t, %, IPI, R_{n-1}, and HE_{n-1}. These interaction terms are specified with the general format of $\beta_{x*n-1}x*n\text{-}1$ where x represents an individual interaction term (e.g., for time in session t interacting with n - 1 duration: $\beta_{t*n-1}t*n\text{-}1$). These same interaction terms were included with the moving average term (MA, of the general format $\beta_{x*MA}x*MA$) in order to see if very recent experience (n - 1) and long-

term experience (MA) were differentially influenced by variables such as time. Interestingly, when examining further n-back interactions, only the interaction between IPI_{n-2} and n - 2 duration survived the BIC selection process, indicating that individual further n-backs were less open to modification by these variables.

In the probabilistic reward experiment, we added a trinary term for if a lever press was unsuccessful (0), successful and rewarded (1), or successful and unrewarded (2), and included interactions between this term and n - 1 as well as the MA. Additionally, we ran all three probability groups together in the model and included indicator variables for which group (25%, 50%, or 75% reward) a mouse belonged to. This allowed us to include a 3-way interaction to determine if the groups differed in how this trinary outcome term interacted with prior press durations (e.g., does the probability of reward influence the presence/magnitude of win-stay behavior?). For the optogenetic inhibition LME models, we included a binary term indicating if a lever press received light delivery (before, during, or after for the three different manipulations) as both a main effect and as an interaction with n - 1 duration and the MA to determine if light reduced the relationship between press n and press n - 1/the MA.

Ca^{2+} Activity Linear Mixed Effect Models

For the M2 and M2-DMS Ca^{2+} activity recordings, we built LME models that sought to predict Ca^{2+} activity given behavior. For this, we used data only from the 1600ms training days. First, we built simple LME models that included only current (n) and prior (n-back, up to n - 6) durations to predict activity (calculated as area under the

curve) at three different time points: -1s to 0s before press onset, during the lever press itself, and 0s to +1s after press release. For activity during the lever press itself, we used modified Akima interpolation, implemented using Matlab's interp1 function to get presses of different durations on the same relative scale, and we excluded any lever presses with fewer than 2 samples which would preclude interpolation. We also included terms for prior activity (up to n - 6) to control for autocorrelation in the Ca^{2+} activity signal. We again compared beta coefficients from the actual data to 1000 order shuffled datasets for these simple models.

$$A = \beta_0 + \beta_n n_0 + \beta_{n-1} n_{-1} + \ldots + \beta_{n-6} n_{-6} + \beta_{A-1} A_{-1} + \ldots + \beta_{A-6} A_{-6} + (1|M) + (1|D) + \varepsilon_i$$

Where A is current Ca^{2+} activity and β_x is the regression coefficient for term x. Of note, these models included n duration (n_0) as a predictor (whereas this was what we predicted in the pure behavioral models). We predicted A given both current (n_0) and prior (n - 1, up to n - 6) press durations, included prior Ca^{2+} activity (A - 1, up to A - 6) as a covariate, and included random intercepts of mouse (M) and training day (D).

Additionally, we built more complex LME models to predict Ca^{2+} activity data. For these, we used the complex behavioral model above for the predictors, as we were interested in seeing if these variables - which we know influence the behavior - are also represented in Ca^{2+} activity, and also still included prior Ca^{2+} activity data to control for autocorrelation in the Ca^{2+} data. This took the form of the following equation, using the same variables as the preceding equations.

$$A = \beta_0 + \beta_{A\text{-}1}A_{\text{-}1} + \ldots + \beta_{A\text{-}6}A_{\text{-}6} + \beta_{n0}n_0 + \beta_{n\text{-}1}n_{\text{-}1} + \ldots + \beta_{n\text{-}6}n_{\text{-}6} + \beta_{MA}MA + \beta_t t + \beta_{\%}\% + \beta_{IPI}IPI_{\text{-}1} + \beta_{IPI}IPI_{\text{-}2} + \beta_R R_{\text{-}1} + \beta_{HE}HE_{\text{-}1} + \beta_{t*n\text{-}1}t*n\text{-}1 + \beta_{\%*n\text{-}1}\%*n\text{-}1 + \beta_{IPI*n\text{-}1}IPI_{\text{-}1}*n\text{-}1 + \beta_{IPI*n\text{-}2}IPI_{\text{-}2}*n\text{-}2 + \beta_{R*n\text{-}1}R_{\text{-}1}*n\text{-}1 + \beta_{HE*n\text{-}1}HE_{\text{-}1}*n\text{-}1 + \beta_{t*MA}t*MA + \beta_{\%*MA}\%*MA + \beta_{IPI*MA}IPI_{\text{-}1}*MA + \beta_{R*MA}R_{\text{-}1}MA + \beta_{HE*MA}HE_{\text{-}1}*MA + (1|M) + (1|D) + \varepsilon_i$$

When trying to predict activity after lever press offset, we also included a binary term for outcome on lever press n (R_0 i.e., the lever press that was just completed with 0 being no reward and 1 being reward). We did not include this term at the other time points since it would introduce a “post diction” confound (i.e., including a term for the outcome of a press before the press even occurred at onset). For the same reason, we did not include interactions with the n_0 variable.

Quantification and Statistical Analysis

All analyses were two-tailed with $\alpha = 0.05$ as a threshold for significance. For analyzing coarse behavioral measurements (e.g., Total Lever Presses) one-way or two-way RM ANOVAs were used, with Greenhouse-Geisser correction for one-way ANOVA and Bonferroni corrections for post-hoc multiple comparisons unless otherwise noted. We used the RMcorr package (Bakdash & Marusich, 2017) implemented in R (R Core Team) to calculate a repeated measures correlation between individual model fit and mouse performance to account for the repeated nature of this data (sampling individual mice across days). We used Matlab’s cusum function to get the upper cumulative sum in Figures 2.1I-J, using 2 SD as the criterion. In our simple LME models, we used permutation tests comparing actual β coefficient values to a distribution of 1000 order shuffled versions of the same variable, and thus

the resolution of our permutation p-values was $p < 0.001$. We excluded presses over 10s in duration from all modeling datasets. For event-aligned Ca^{2+} activity comparing Met vs. Fail lever presses, we used permutation tests that required either 4 (for onset and offset-aligned activity) or 3 (for interpolated activity during the press) consecutive samples passed the threshold for significance. To assess the relationship between Ca^{2+} activity and various aspects of behavior in our complex LME models, we performed F-tests on the individual parameters. For group comparisons (e.g., Sham vs. Lesion) of LME model coefficients, we used t-tests with Benjamini-Hochberg false discovery rate correction (Q = 5%) on all of the model terms shown in Tables 2.S4 and 2.S7. Behavioral data was analyzed using Excel (Microsoft), Matlab (Mathworks), R (R Core Team), and Prism (Graphpad).

Data Availability

The data reported in this paper will be shared by the lead contact upon request.

Code Availability

The code used to analyze the data from this study is available at: https://github.com/gremellab/Hold-Down-Behavior-GCAMP-Opto-analysis

Author Contributions

D.C.S.: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing - original draft, Writing - review and editing. C.C.: Formal analysis and Writing - review and editing. R.R.: Investigation and Writing -

review and editing. C.M.G.: Conceptualization, Supervision, Funding acquisition, Project administration, Visualization, Writing - original draft, Writing - review and editing.

Declaration of Interests

The authors declare no competing interests.

Supplemental Information

Note 2.S1, related to Figure 2.1. As previous reports have indicated that rats in a similarly unstructured task may sometimes make short, stereotyped lever presses after reward (Platt et al., 1973), we investigated our data for evidence of stereotypies. Of the 13 behavioral mice, we found evidence of 1 mouse that appeared to adopt this stereotyped strategy, making presses after a reward that were 450 ± 434ms (mean ± SD) in duration, while the average for all other animals was 1051 ± 757ms. Permutation tests comparing to order shuffled data found that this same mouse exhibited a smaller SD after a rewarded press in actual versus shuffled data on 7 out of 14 days, while no other mouse in this (or subsequent) experiments did so for more than 2 days. Thus, while it is possible for animals to adopt a stereotyped strategy to perform this task, only a very small minority of animals appear to do so. Aside from the difference in species (rats versus mice), the results of Platt et al. (1973) may have been due to the extensive 2 week pretraining period without a duration requirement, wherein rats would have been incentivized to press as rapidly as possible to earn

maximal reward and might develop a habitual or stereotyped response that persisted even after introduction of the duration requirement.

Acknowledgements

Chapter 2 is, in full, a reprint of the material submitted to *Nature Neuroscience*. Schreiner, Drew C.; Cazares, Christian; Renteria, Rafael; Gremel, Christina M. The dissertation author was the primary investigator and author of this paper. This project was funded by the NIH: F31AA027439 (D.C.S.), F32AA026776 (R.R.), R01AA026077 (C.M.G.), the NSF: DGE-1650112 (C.C.), and a Whitehall Foundation Award (C.M.G).

References

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, *33*(2), 109–121. https://doi.org/10.1080/14640748108400816

Allain, F., Minogianis, E.-A., Roberts, D. C. S., & Samaha, A.-N. (2015). How fast and how often: The pharmacokinetics of drug use are decisive in addiction. *Neuroscience & Biobehavioral Reviews*, *56*, 166–179. https://doi.org/10.1016/j.neubiorev.2015.06.012

Allen, W. E., Kauvar, I. V., Chen, M. Z., Richman, E. B., Yang, S. J., Chan, K., Gradinaru, V., Deverman, B. E., Luo, L., & Deisseroth, K. (2017). Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron*, *94*(4), 891-907.e6. https://doi.org/10.1016/j.neuron.2017.04.017

Ariely, D., & Zakay, D. (2001). A timely account of the role of duration in decision making. *Acta Psychologica*, *108*(2), 187–207. https://doi.org/10.1016/S0001-6918(01)00034-8

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00456

Balleine, B. W. (2019). The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. *Neuron*, *104*(1), 47–62. https://doi.org/10.1016/j.neuron.2019.09.024

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4–5), 407–419. https://doi.org/10.1016/S0028-3908(98)00033-1

Berridge, K. C., Zhang, J., & Aldridge, J. W. (2008). Computing motivation: Incentive salience boosts of drug or appetite states. *The Behavioral and Brain Sciences*, *31*(4), 440–441. https://doi.org/10.1017/S0140525X08004767

Bouton, M. E., & Balleine, B. W. (2019). Prediction and control of operant behavior: What you see is not all there is. *Behavior Analysis: Research and Practice*, *19*(2), 202–212. https://doi.org/10.1037/bar0000108

Busse, L., Ayaz, A., Dhruv, N. T., Katzner, S., Saleem, A. B., Schölvinck, M. L., Zaharia, A. D., & Carandini, M. (2011). The Detection of Visual Contrast in the Behaving Mouse. *The Journal of Neuroscience*, *31*(31), 11351–11361. https://doi.org/10.1523/JNEUROSCI.6689-10.2011

Corbit, V. L., Manning, E. E., Gittis, A. H., & Ahmari, S. E. (2019). Strengthened Inputs from Secondary Motor Cortex to Striatum in a Mouse Model of Compulsive

Behavior. *Journal of Neuroscience*, *39*(15), 2965–2975. https://doi.org/10.1523/JNEUROSCI.1728-18.2018

Costa, R. M. (2011). A selectionist account of de novo action learning. *Current Opinion in Neurobiology*, *21*(4), 579–586. https://doi.org/10.1016/j.conb.2011.05.004

Cristol, D. A., & Switzer, P. V. (1999). Avian prey-dropping behavior. II. American crows and walnuts. *Behavioral Ecology*, *10*(3), 220–226. https://doi.org/10.1093/beheco/10.3.220

Delevich, K., Okada, N. J., Rahane, A., Zhang, Z., Hall, C. D., & Wilbrecht, L. (2020). Sex and Pubertal Status Influence Dendritic Spine Density on Frontal Corticostriatal Projection Neurons in Mice. *Cerebral Cortex*, *30*(6), 3543–3557. https://doi.org/10.1093/cercor/bhz325

Dhawale, A. K., Miyamoto, Y. R., Smith, M. A., & Ölveczky, B. P. (2019). Adaptive Regulation of Motor Variability. *Current Biology*, *29*(21), 3551-3562.e7. https://doi.org/10.1016/j.cub.2019.08.052

Dickinson, A. (1985). Actions and Habits: The Development of Behavioural Autonomy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *308*(1135), 67–78. https://doi.org/10.1098/rstb.1985.0010

Ebbesen, C. L., Insanally, M. N., Kopec, C. D., Murakami, M., Saiki, A., & Erlich, J. C. (2018). More than Just a “Motor”: Recent Surprises from the Frontal Cortex. *Journal of Neuroscience*, *38*(44), 9402–9413. https://doi.org/10.1523/JNEUROSCI.1671-18.2018

Erlich, J. C., Bialek, M., & Brody, C. D. (2011). A Cortical Substrate for Memory-Guided Orienting in the Rat. *Neuron*, *72*(2), 330–343. https://doi.org/10.1016/j.neuron.2011.07.010

Fan, D., Rossi, M. A., & Yin, H. H. (2012). Mechanisms of Action Selection and Timing in Substantia Nigra Neurons. *Journal of Neuroscience*, *32*(16), 5534–5548. https://doi.org/10.1523/JNEUROSCI.5924-11.2012

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. *Annals of the New York Academy of Sciences*, *423*(1), 52–77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x

Gomes, P. V. O., Brasil-Neto, J. P., Allam, N., & Rodrigues de Souza, E. (2012). A Randomized, Double-Blind Trial of Repetitive Transcranial Magnetic Stimulation in Obsessive-Compulsive Disorder With Three-Month Follow-Up. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *24*(4), 437–443. https://doi.org/10.1176/appi.neuropsych.11100242

Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M., & Mainen, Z. F. (2014). Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, *17*(11), 1455–1462. https://doi.org/10.1038/nn.3812

Gorzelak, M. A., Asay, A. K., Pickles, B. J., & Simard, S. W. (2015). Inter-plant communication through mycorrhizal networks mediates complex adaptive behaviour in plant communities. *AoB PLANTS*, *7*. https://doi.org/10.1093/aobpla/plv050

Gremel, C. M., & Costa, R. M. (2013). Premotor cortex is critical for goal-directed actions. *Frontiers in Computational Neuroscience*, *7*. https://doi.org/10.3389/fncom.2013.00110

Hawken, E. R., Dilkov, D., Kaludiev, E., Simek, S., Zhang, F., & Milev, R. (2016). Transcranial Magnetic Stimulation of the Supplementary Motor Area in the Treatment of Obsessive-Compulsive Disorder: A Multi-Site Study. *International Journal of Molecular Sciences*, *17*(3). https://doi.org/10.3390/ijms17030420

Hintiryan, H., Foster, N. N., Bowman, I., Bay, M., Song, M. Y., Gou, L., Yamashita, S., Bienkowski, M. S., Zingg, B., Zhu, M., Yang, X. W., Shih, J. C., Toga, A. W., & Dong, H.-W. (2016). The mouse cortico-striatal projectome. *Nature Neuroscience*, *19*(8), 1100–1114. https://doi.org/10.1038/nn.4332

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), 1–10. https://doi.org/10.1038/s41467-018-04840-2

Jean-Richard-dit-Bressel, P., Clifford, C. W. G., & McNally, G. P. (2020). Analyzing Event-Related Transients: Confidence Intervals, Permutation Tests, and Consecutive Thresholds. *Frontiers in Molecular Neuroscience*, *13*. https://doi.org/10.3389/fnmol.2020.00014

Johnson, C. M., Peckler, H., Tai, L.-H., & Wilbrecht, L. (2016). Rule learning enhances structural plasticity of long-range axons in frontal cortex. *Nature Communications*, *7*, 10785. https://doi.org/10.1038/ncomms10785

Juavinett, A. L., Erlich, J. C., & Churchland, A. K. (2018). Decision-making behaviors: Weighing ethology, complexity, and sensorimotor compatibility. *Current Opinion in Neurobiology*, *49*, 42–50. https://doi.org/10.1016/j.conb.2017.11.001

Klaus, A., da Silva, J. A., & Costa, R. M. (2019). What, If, and When to Move: Basal Ganglia Circuits and Self-Paced Action Initiation. *Annual Review of Neuroscience*, *42*(1), null. https://doi.org/10.1146/annurev-neuro-072116-031033

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Lak, A., Hueske, E., Hirokawa, J., Masset, P., Ott, T., Urai, A. E., Donner, T. H., Carandini, M., Tonegawa, S., Uchida, N., & Kepecs, A. (2020). Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. *ELife*, *9*, e49834. https://doi.org/10.7554/eLife.49834

Lopes, G., Bonacchi, N., Frazão, J., Neto, J. P., Atallah, B. V., Soares, S., Moreira, L., Matias, S., Itskov, P. M., Correia, P. A., Medina, R. E., Calcaterra, L., Dreosti, E., Paton, J. J., & Kampff, A. R. (2015). Bonsai: An event-based framework for processing and controlling data streams. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00007

Lünemann, A., Ullrich, O., Diestel, A., Jöns, T., Ninnemann, O., Kovac, A., Pohl, E. E., Hass, R., Nitsch, R., & Hendrix, S. (2006). Macrophage/microglia activation factor expression is restricted to lesion-associated microglial cells after brain trauma. *Glia*, *53*(4), 412–419. https://doi.org/10.1002/glia.20301

Mantovani, A., Rossi, S., Bassi, B. D., Simpson, H. B., Fallon, B. A., & Lisanby, S. H. (2013). Modulation of motor cortex excitability in obsessive-compulsive disorder: An exploratory study on the relations of neurophysiology measures with clinical outcome. *Psychiatry Research*, *210*(3), 1026–1032. https://doi.org/10.1016/j.psychres.2013.08.054

McDougle, S. D., Butcher, P. A., Parvin, D. E., Mushtaq, F., Niv, Y., Ivry, R. B., & Taylor, J. A. (2019). Neural Signatures of Prediction Errors in a Decision-Making Task Are Modulated by Action Execution Failures. *Current Biology : CB*, *29*(10), 1606-1613.e5. https://doi.org/10.1016/j.cub.2019.04.011

Mooney, R. (2009). Neural mechanisms for learned birdsong. *Learning & Memory*, *16*(11), 655–669. https://doi.org/10.1101/lm.1065209

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

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

Pinto, L., Rajan, K., DePasquale, B., Thiberge, S. Y., Tank, D. W., & Brody, C. D. (2019). Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions. *Neuron*, *104*(4), 810-824.e9. https://doi.org/10.1016/j.neuron.2019.08.025

Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., & Churchland, A. K. (2021). Lapses in perceptual decisions reflect exploration. *ELife*, *10*, e55490. https://doi.org/10.7554/eLife.55490

Platt, J. R., Kuch, D. O., & Bitgood, S. C. (1973). Rats' lever-press durations as psychophysical judgments of time. *Journal of the Experimental Analysis of Behavior*, *19*(2), 239–250. https://doi.org/10.1901/jeab.1973.19-239

Remedios, R., Kennedy, A., Zelikowsky, M., Grewe, B. F., Schnitzer, M. J., & Anderson, D. J. (2017). Social behaviour shapes hypothalamic neural ensemble representations of conspecific sex. *Nature*, *550*(7676), 388–392. https://doi.org/10.1038/nature23885

Rescorla, R. A., & Wagner, A. R. (1972). *3 A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement*. 18.

Rothwell, P. E., Hayton, S. J., Sun, G. L., Fuccillo, M. V., Lim, B. K., & Malenka, R. C. (2015). Input- and Output-Specific Regulation of Serial Order Performance by Corticostriatal Circuits. *Neuron*, *88*(2), 345–356. https://doi.org/10.1016/j.neuron.2015.09.035

Roy, N. A., Bak, J. H., Akrami, A., Brody, C. D., & Pillow, J. W. (2021). Extracting the dynamics of behavior in sensory decision-making experiments. *Neuron*, *109*(4), 597-610.e6. https://doi.org/10.1016/j.neuron.2020.12.004

Schreiner, D. C., & Gremel, C. M. (2018). Orbital Frontal Cortex Projections to Secondary Motor Cortex Mediate Exploitation of Learned Rules. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-29285-x

Schreiner, D. C., Yalcinbas, E. A., & Gremel, C. M. (2021). A push for examining subjective experience in value-based decision-making. *Current Opinion in Behavioral Sciences*, *41*, 45–49. https://doi.org/10.1016/j.cobeha.2021.03.020

Setsuie, R., Tamura, K., Miyamoto, K., Watanabe, T., Takeda, M., & Miyashita, Y. (2020). Off-Peak 594-nm Light Surpasses On-Peak 532-nm Light in Silencing Distant ArchT-Expressing Neurons In Vivo. *IScience*, *23*(7), 101276. https://doi.org/10.1016/j.isci.2020.101276

Shimaoka, D., Harris, K. D., & Carandini, M. (2018). Effects of Arousal on Mouse Sensory Cortex Depend on Modality. *Cell Reports*, *22*(12), 3160–3167. https://doi.org/10.1016/j.celrep.2018.02.092

Siniscalchi, M. J., Phoumthipphavong, V., Ali, F., Lozano, M., & Kwan, A. C. (2016). Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior. *Nature Neuroscience*, *19*(9), 1234–1242. https://doi.org/10.1038/nn.4342

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis* (p. 457). Appleton-Century.

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 1–8. https://doi.org/10.1038/s41586-019-1787-x

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. 551.

Tervo, D. G. R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., & Karpova, A. Y. (2014). Behavioral Variability through Stochastic Choice and Its Gating by Anterior Cingulate Cortex. *Cell*, *159*(1), 21–32. https://doi.org/10.1016/j.cell.2014.08.037

Yin, H. H. (2009). The role of the murine motor cortex in action duration and order. *Frontiers in Integrative Neuroscience*, *3*, 23. https://doi.org/10.3389/neuro.07.023.2009

Yoo, S. B. M., Hayden, B. Y., & Pearson, J. M. (2021). Continuous decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1819), 20190664. https://doi.org/10.1098/rstb.2019.0664

CHAPTER THREE

Chronic alcohol exposure disrupts behavioral flexibility via hyperactive premotor corticostriatal activity

Drew C. Schreiner[1], Rafael Renteria[1], Emily T. Baltz[2], Christian Cazares[2], and Christina M. Gremel[1,2]

[1]Department of Psychology, University of California San Diego, La Jolla, California, 92093

[2]Neurosciences Graduate Program, University of California, San Diego, La Jolla, California, 92093

Abstract

Background

Alcohol Use Disorder (AUD) disrupts behavioral flexibility, but alterations to underlying neural circuit mechanisms are unclear. One candidate is the premotor corticostriatal circuit which has been implicated in compulsive and inflexible behaviors and shows altered activity in AUD. However, whether AUD alters mechanisms within premotor corticostriatal circuits to disrupt behavioral flexibility is unknown.

Methods

Male and female C57BL/6J mice underwent chronic intermittent ethanol (CIE) or Air vapor exposure and repeated withdrawal. Mice were subsequently trained on an instrumental lever hold-down task, which allowed for modeling the contribution of recent experience important for behavioral flexibility. *Ex vivo* slice recordings examined effects of vapor exposure on intrinsic properties of premotor cortex (M2). *In vivo* calcium fiber photometry in vapor-exposed mice examined activity and modulation of M2 neurons with projections into dorsal medial striatum (M2-DMS) during task performance. Finally, a projection-specific inhibitory chemogenetic approach was used to examine a causal link between CIE disruption to both M2-DMS activity and behavioral flexibility.

Results

Prior CIE impaired behavioral flexibility and was accompanied by a CIE-induced increase in M2 excitability as well as increased *in vivo* activity of M2-DMS projection

neurons at baseline and during task performance. Chemogenetic reduction of this CIE-induced hyperactivity in M2-DMS neurons rescued behavioral flexibility.

Conclusions

This is the first evidence of a direct, causal relationship between chronic alcohol disruption to premotor circuits and behavioral flexibility and provides mechanistic support for targeting activity of human premotor regions as a potential treatment in AUD.

Introduction

Alcohol Use Disorder (AUD) is associated with disruptions to behavioral flexibility (Claus et al., 2011; Duka et al., 2011; Scaife & Duka, 2009; Shnitko et al., 2020; Sjoerds et al., 2014), defined as the ability to appropriately adapt behavior based on changing circumstances. These disruptions contribute to daily dysfunction, continued alcohol abuse, and relapse (Belin et al., 2013; Everitt & Robbins, 2005; Hogarth, 2020), suggesting that targeting restoration of function has treatment potential (Gremel & Lovinger, 2017). Preclinical work has found AUD-related alterations to regions important for behavioral flexibility, including cortex (den Hartog et al., 2016; Morningstar et al., 2020; Nimitvilai et al., 2016, 2017; Renteria et al., 2018, 2021) and cortical output into dorsal striatum (Carlson, 2018; Carlson et al., 2011; Lovinger & Alvarez, 2017; Ma et al., 2018; Muñoz et al., 2018; Patton et al., 2021; Renteria et al., 2018, 2021), the main input nucleus of the basal ganglia. However, the specific circuits and mechanisms disrupted in AUD that result in behavioral inflexibility are not clear. As novel strategies are being explored in the treatment of AUD, including rTMS targeting of dorsal cortex, it is essential to identify circuit-specific mechanisms involved.

One potential but understudied candidate is the primate pre-supplementary and supplementary motor areas (Pre-SMA/SMA) which send projections into the caudate nucleus and are disrupted in AUD (Claus et al., 2011; Duka et al., 2011; Morris et al., 2018; Sjoerds et al., 2014). Pre-SMA/SMA are involved in supporting behavioral flexibility (Aron, 2011; Aron et al., 2007; Morris, Kundu, et al., 2016), with altered

activity implicated in compulsive disease states (Gomes et al., 2012; Hawken et al., 2016; Mantovani et al., 2013). Correlative studies have shown that SMA activity positively correlates with AUD severity, with more severe symptoms associated with greater recruitment of SMA activity during delayed discounting (Claus et al., 2011). In addition, abstinent individuals with AUD show impaired response inhibition and reduced SMA volume (Duka et al., 2011) and reduced activity in a response inhibition task (Sjoerds et al., 2014). However, missing is a causal link between AUD effects on Pre-SMA/SMA function and impaired flexibility.

Limited preclinical work has examined Pre-SMA/SMA's rodent homologue (Barthas & Kwan, 2017), the premotor cortex (M2, also known as secondary motor cortex) in relation to AUD. Research has shown M2 activity represents information important for behavioral flexibility (e.g., prior as well as future actions), and is modulated by experience (Hattori et al., 2019; Murakami et al., 2014, 2017; Siniscalchi et al., 2019). M2 sends dense projections to the dorsal medial striatum (DMS, akin to primate caudate nucleus) (Delevich et al., 2020; Hintiryan et al., 2016). This M2-DMS projection is strengthened in Obsessive Compulsive Disorder (V. L. Corbit et al., 2019), and is involved in working memory deficits associated with Parkinson's disease (Magno et al., 2019) in rodent models. However, only limited examination of M2 has been done in the context of alcohol. Brain-wide scans revealed acute alcohol increased cFos within M2 (Liu & Crews, 2015) and chronic alcohol increased M2 activity assessed via MRI (Dudek et al., 2015). This suggests chronic alcohol exposure may affect the activity and function of M2, and may alter its contribution to behavioral flexibility.

Here we examined whether chronic alcohol induced long-lasting changes to M2 activity and function. We used a well-validated model of chronic alcohol exposure (Becker & Hale, 1993; Becker & Lopez, 2004, 2006; Cazares et al., 2021; Lopez & Becker, 2005; Renteria et al., 2018), and examined neural activity and behavior in protracted withdrawal (Heilig et al., 2010). Of note, we employed an instrumental task (Cazares et al., 2021; Fan et al., 2012; Platt et al., 1973; Yin, 2009) that allowed for the continuous analysis of behavioral flexibility from one decision to the next and online investigation of associated neural mechanisms. Use of this task bypassed limitations often present in traditional reversal learning, set-shifting, or response inhibition investigations which may occlude examination of how recent experience is used to flexibly control behavior (e.g., often averaging a limited number of discrete responses across time (Schreiner et al., 2021)). Thus, we were able to examine persistent chronic alcohol effects on M2 activity, and its contribution to behavioral flexibility in a continuous manner.

Methods and Materials

Experimental Model and Subject Details

Similar numbers of male and female C57BL/6J mice (>7 weeks/50 PND) (The Jackson Laboratory, Bar Harbor, ME) were used for all experiments. As exploratory analyses for sex differences in the behavioral cohort (50% M/F split) revealed no differences, data were collapsed across sex. All procedures were conducted during the light period and mice had free access to water. Mice were housed 2–4 per cage on a 14:10 light:dark cycle. Mice were food restricted to 85-90% of their baseline weight 2

days prior to starting behavioral procedures, and fed daily 1–4 hours after training. All experiments were approved by the University of California San Diego Institutional Animal Care and Use Committee and were carried out in accordance with the National Institutes of Health (NIH) “Principles of Laboratory Care”.

Surgical Procedures

Viral vectors were obtained from the UNC Viral Vector Core (Chapel Hill, NC) or Addgene (Watertown, MA). Mice were anesthetized with isoflurane (1-2%) and intracranial injections (100 nl/min) were targeted at a relative posterior portion of M2 (from Bregma: AP +1.0mm, L ±0.5mm and V -1.2mm, -1.4mm from the skull), and DMS (from Bregma: AP +1.0mm, L ±1.65mm and V -3.0mm, -3.2mm from the skull). Mice were given at least one week of recovery prior to the start of CIE procedures. After behavioral testing was completed, mice were euthanized and brains were extracted and fixed in 4% paraformaldehyde. Virus localization and spread was assessed in 100 µm thick brain slices via fluorescent microscopy (Olympus MVX10, Shinjuku, Japan).

For M2-DMS calcium imaging, n = 8 AIR and n = 8 CIE mice (4 M/F per group) were unilaterally injected with a virus expressing a cre-dependent GCaMP6s in M2 (pAAV.CAG.FLEX.GCaMP6s.WPRE.SV40 (Addgene: 100842): 2 injection depths, 200nl each), and a retrograde-capable virus (Rothermel et al., 2013) expressing cre recombinase (AAV5/Ef1a-Cre-WPRE) in DMS (2 injection depths, 250nl each). Animals were then implanted with an optical ferrule centered on M2. Two AIR mice were excluded due to poor viral expression (to give final n = 6 AIR).

For chemogenetic inhibition of M2-DMS, control animals (n = 6 AIR (4/2 MF), n = 6 CIE (3/3 M/F) were injected with a virus expressing cre-dependent tdTomato in M2 (rAAV5/Flex-tdTomato), while experimental animals (n = 9 AIR, n = 8 CIE) were injected with a virus expressing a cre-dependent inhibitory chemogenetic receptor hM4Di (H4) in M2 (pAAV5-hSyn-DIO-hM4D(Gi)-mCherry: 2 injection depths, 150nl each). Both groups were injected with a retrograde-capable virus (Rothermel et al., 2013) expressing cre recombinase in DMS (rAAV5/hSyn-GFP-Cre: 2 injection depths, 200nl each). One AIR H4 mouse, and one CIE H4 mouse were excluded due to viral expression (final n = 8 AIR H4 (4/4 M/F) and n = 7 CIE H4 (4/3 M/F)).

Chronic Intermittent Ethanol Exposure and Withdrawal

Mice (>8 weeks of age) were exposed to chronic intermittent ethanol vapor (CIE) or Air (Becker & Hale, 1993; Becker & Lopez, 2004, 2006; Cazares et al., 2021; Lopez & Becker, 2005; Renteria et al., 2018). As previously described (Cazares et al., 2021; Renteria et al., 2018, 2021), mice were exposed to Air/CIE vapor for 16 hrs/day, for four consecutive days, and this procedure was repeated for 4 weeks. Ethanol was volatilized by bubbling air through a flask containing 95% ethanol at a rate of 2/3 L/min, delivered to the mice housed in Plexiglas containers (Plas Labs Inc.). No loading dose of ethanol or pyrazole pretreatment was administered to avoid confounding effects of 1) stress on behavior (Dias-Ferreira et al., 2009) and 2) pyrazole on neural activity (Becker & Lopez, 2004; Pereira et al., 1992). Blood alcohol concentration was collected from sentinel animals, with a mean±SEM of 27.9 ± 2.0mM.

Behavioral Procedures

To examine CIE-effects that persist into protracted withdrawal (Heilig et al., 2010), five days post the final vapor exposure, mice began daily operant training in sound-attenuating boxes (Med-Associates, St Albans, VT) in which they pressed a lever (left or right of the food magazine, counterbalanced) for an outcome of regular 'chow' pellets (20 mg pellet per reinforcer, Bio-Serv formula F0071). Each training session commenced with an illumination of the house light and lever extension and ended after either 60 reinforcers were earned or 90 minutes had elapsed, with the lever retracting and the house light turning off.

On the first day, mice were trained to approach the food magazine to retrieve the pellet outcome (no lever present) on a random time (RT 120s) schedule, for a total of 60 minutes. Next, mice were trained on a continuous ratio schedule of reinforcement (CRF) across 3 days, where every lever press was reinforced, and the total possible number of reinforcers increased (CRF10, 30, and 60) across days.

Following CRF pretraining, the lever-press duration contingency was introduced, in which mice had to press and hold down the lever for a minimum duration in order to earn food reward (delivered immediately after press release). Importantly, there were no cues, no timeout period, nor any discrete trials; the lever was always available to mice, until the session was complete. Mice were trained at the >800ms criterion for 6 days, followed by at least 6 days of training at the >1600ms criterion. Timestamps for lever press onset and offset, headentry into the food magazine onset and offset, and reward delivery were recorded. From these

timestamps, we calculated durations of lever presses and headentries (limit of 20ms resolution).

Fiber Photometry

Animals received one additional day of CRF pretraining when they were first connected to 400 um optical fiber tethers with ferrule-to-ferrule connectivity. GCaMP6s was excited using a 470nm LED at < 70 µW/mm2 (Thorlabs, Newton, NJ). GCaMP6s fluorescence emission was collected using a bifurcated fiber (Thorlabs, Newton, NJ) that allowed for simultaneous, independent recordings from two mice. The dual fiber core was imaged using a 4X objective (Olympus) focused onto a CMOS camera (FLIR Systems, Wilsonville, OR). Bonsai software (Lopes et al., 2015) was used to select the fiber cores and acquire fluorescence intensity signals at 20 Hz. Bonsai simultaneously collected analog timestamps for lever presses, headentries, reward delivery via TTL pulses sent from MED-PC and collected using Arduino Duo microprocessors (Arduino, Somerville, MA) with custom code. Photometry and behavioral data were imported into Matlab (Mathworks Inc., Natick, MA) for analysis using custom scripts (https://github.com/gremellab/Hold-Down-Behavior-GCAMP-Opto-analysis). We fit the fluorescence intensity signal to a double exponential to account for photobleaching across a session. To check for bad coupling of the fiber to the ferrule or low expression, each session we calculated the 97.5 percentile of dF/F and ensured that there was at least a 1% change; sessions failing to meet this criterion were excluded from analyses (Markowitz et al., 2018). We also excluded sessions with visual anomalies in the session long traces (e.g., a sudden, sustained jump in activity that

could indicate fiber decoupling). For calcium transient analyses (Figures 3.3B-C), we used Matlab's findpeaks function, with the "MinPeakHeight" argument set to the 4*median absolute deviation plus the mean of the session-long calcium signal (Pribiag et al., 2021). We used the mean and standard deviation during a baseline period -15s to -5s preceding the lever press to z-score press-related activity (i.e., from -5s prior to onset up to 5s post offset). The session long mean was used to calculate DF/F as: ((F - Fmean)/Fmean) x 100%. We bootstrapped 99% confidence intervals using the boot_CI function (Jean-Richard-dit-Bressel et al., 2020). For examining activity during lever pressing, we used Makima interpolation with Matlab's interp1 function. To compare activity in AIR and CIE mice, we performed running permutation tests that required at least 4 adjacent samples to significantly differ from one another (Jean-Richard-dit-Bressel et al., 2020). Calcium activity data was smoothed using a 10 sample (or 5 sample for interpolated activity) long Gaussian filter for display purposes only.

Chemogenetic Inhibition

Animals underwent behavioral training as above. To target behavioral measurements to a time period overlapping with circuit attenuation (Gremel & Costa, 2013, p. 2) that also avoids indirect effects of agonist treatment (Gomez et al., 2017), all mice were given intraperitoneal injections of the hMm4Di selective agonist Clozapine-N-Oxide (CNO, 1.0 mg/kg, 10 ml/kg, Sigma Aldrich) 30 minutes prior to each hold down training session.

Slice Electrophysiology

Coronal slices (250 μm thick) containing M2 were prepared using a Pelco easiSlicer (Ted Pella Inc, Redding, CA). Mice were anesthetized by isoflurane and brains were rapidly removed and placed in 4°C oxygenated ACSF containing the following (in mM): 210 sucrose, 26.2 $NaHCO_3$, 1 NaH_2PO_4, 2.5 KCl, 11 dextrose, bubbled with 95% O_2/5% CO_2. Slices were transferred to an ACSF solution for incubation containing the following (in mM): 120 NaCl, 25 $NaHCO_3$, 1.23 NaH_2PO_4, 3.3 KCl, 2.4 $MgCl_2$, 1.8 $CaCl_2$, 10 dextrose. Slices were continuously bubbled with 95% O_2/5% CO_2 at pH 7.4, 32°C and maintained in this solution for at least 60 min prior to recording. Whole-cell current clamp recordings were made in M2 pyramidal cells. Pyramidal cells were identified using an Olympus BX51WI microscope mounted on a vibration isolation table. Recordings were made in ACSF containing (in mM): 120 NaCl, 25 $NaHCO_3$, 1.23 NaH_2PO_4, 3.3 KCl, 0.9 $MgCl_2$, 2.0 $CaCl_2$, and 10 dextrose, bubbled with 95% O_2/5% CO_2. ACSF was continuously perfused at a rate of 2.0 mL/min and maintained at a temperature of 32°C. Picrotoxin (50 μM) was included in the recording ACSF to block $GABA_A$ receptor-mediated synaptic currents. Recording electrodes (thin-wall glass, WPI Instruments) were made using a PC-10 puller (Narishige International, Amityville, NY) to yield resistances between 3–6 MΩ. Electrodes were filled with (in mM): 135 $KMeSO_4$, 12 NaCl, 0.5 EGTA, 10 HEPES, 2 Mg-ATP, 0.3 Tris-GTP, 260–270 mOsm (pH 7.3). Access resistance was monitored throughout the experiments, and cells in which it varied more than 20% were excluded. Recordings were made using a MultiClamp 700B amplifier (Molecular Devices, Union City, CA), filtered at 2 kHz, digitized at 10 kHz with Instrutech ITC-18 (HEKA Instruments, Bellmore, NY), and displayed and saved using AxographX

(Axograph, Sydney, Australia). A series of fixed current injections (20 pA increments from 0 to 240 pA) were used to elicit action potential firing and the number of spikes were counted at each current step. For CNO verification, we used a concentration of 10μM CNO in recording ACSF as described above. First, we took a baseline measure using the series of fixed current injections. CNO was washed on for 5 minutes prior to recording and again, a series of fixed current injections were used to elicit action potential firing. CNO was washed off for 5 minutes and a baseline was recorded again.

Data Analysis

Linear Mixed Effect Models

We built Linear Mixed Effects models (LME) to model the relationship between the duration of lever press n and n-back (n - 1 through n - 10) lever press durations. We included random intercept terms for mouse and day of training to account for the repeated structure of our data. To determine how far back a significant relationship existed between press n and any particular n-back press, we shuffled the order of a particular n-back (e.g., only n - 3) 1000 times and compared the shuffled distribution of beta coefficients to the actual value via permutation test. Of note, we are shuffling here within individual mouse/sessions, thus preserving the overall statistics of the data, and shuffling only the order in which a specific type of event occurred.

$$n = \beta_0 + \beta_{n-1}n_{-1} + \beta_{n-2}n_{-2} + \ldots + \beta_{n-10}n_{-10} + \beta_t(t) + \beta_{\%}(\%) + (1|M) + (1|D) + \varepsilon_i$$

Where n is the current lever press duration, n - 1 through n - 10 are the previous 1 through 10 lever press durations and β_x is the linear regression coefficient

for term x (β_0 is the intercept term). We also included covariates of time in session (t) and the percentage of presses that met criteria (%). We included random intercept terms for both mouse (M) and day (D).

Decoding

We used a support vector machine (SVM) classifier trained on individual mouse/sessions to predict press durations using M2-DMS calcium activity data. We created four equal sample duration bins within each individual mouse/session - thus, chance performance was 25%. Next, we used several calcium activity measurements as predictors including the area under the curve, and the slope of the calcium signal from: -1s to 0s prior to press onset, during the press, 0s to +1s after press release, and from +2s to +5s after press release. We then trained the SVM classifier using Matlab's fitcoec function, using additional arguments to standardize the calcium activity data and to specify a linear kernel function. We performed 10 k-fold cross validations on the model, subtracted the classification loss from 1, and multiplied by 100 to get the classification accuracy %.

Quantification and Statistical Analysis

All analyses were two-tailed with $\alpha = 0.05$ as a threshold for significance. For analyzing coarse behavioral measurements (e.g., Total Lever Presses) Repeated Measures (RM) ANOVAs were used (with Greenhouse-Geisser corrections if preliminary analyses indicated different sample standard deviations), with Bonferroni corrections for post-hoc multiple comparisons. We used Mann-Whitney tests when preliminary tests indicated non-normal distributions. In our LME models, we used

permutation tests comparing actual β coefficient values to a distribution of 1000 order shuffled versions of the same variable, and thus the resolution of our permutation p-values was $p < 0.001$. We excluded anomalous presses (>16s in duration) from all datasets. For group comparisons (e.g., AIR vs. CIE) of LME model coefficients, we used 2-way ANOVA, with follow-up post hoc corrected comparisons for individual n-backs. Behavioral data was analyzed using Excel (Microsoft), Matlab (Mathworks), R (R Core Team) and the rmcorr package (Bakdash & Marusich, 2017), JASP, and Prism (Graphpad).

Results

Prior CIE impairs the use of recent experience to guide flexible behavior

The same number of M/F C57BL/6J mice were exposed to 4 weeks of chronic intermittent ethanol exposure and withdrawal (CIE), or air vapor control (AIR) (Figure 3.1A) (Becker & Hale, 1993; Becker & Lopez, 2004, 2006; Cazares et al., 2021; Lopez & Becker, 2005; Renteria et al., 2018). During protracted withdrawal (Heilig et al., 2010), mice were trained to press and hold down a lever for at least a minimum duration in order to earn a food reward (Figure 3.1B). There were no cues or trials and reward was delivered when mice released the lever. In this fully self-initiated and self-paced task, mice were left to rely on their prior experience to guide performance (Figure 3.1C). Thus, our paradigm allowed us to ask if and how prior CIE altered the use of recent experience supporting behavioral flexibility.

We first assessed coarse measurements of behavior and found prior CIE exposure did not affect the number of lever-presses executed (Figure 3.1D) nor the

percentage of presses that met the criterion (Figure 3.1E). In both cases, there was only a main effect of training day, and no Air/CIE group differences. CIE also did not induce a sustained difference in the rate of lever pressing overall, or in the number, or rate of presses that met criteria (Figures 3.S1A-C). Further, the distribution of press durations on the final 800ms and final 1600ms day of training (Figure 3.1F) did not differ between Air and CIE mice, and all mice significantly shifted the distribution of their press durations based on the criterion duration. Collectively, this data suggests that Air and CIE mice were able to acquire and perform this task to relatively similar levels. However, Air and CIE mice could use different behavioral strategies to reach similar levels of performance (e.g. the relative degree of goal-directed/habitual control, or exploration/exploitation, both of which AUD is known to affect (Morris, Baek, et al., 2016; Renteria et al., 2018, 2021)).

One possibility is that Air and CIE mice could rely on prior experience to varying degrees. For instance, CIE may lead to mice making lever press durations that are more or less related to prior durations. To address this question, we built linear mixed effect models (LMEs) to predict the duration of each lever press (n) given the durations of prior lever presses (n - 1 through n - 10). In Figure 3.1G, we report the β coefficients of the individual n-back press covariates in our model. The data shows Air mice relied on the durations of their recent lever presses to inform their current press duration, with the contribution of such experience rapidly decaying across n-back presses. However, reliance on this recent experience was attenuated following CIE exposure, as the lever press durations in CIE-exposed were less related to the duration of the immediately prior lever press. Furthermore, LMEs built using individual session data

showed that only in Air mice did the magnitude of the n - 1 β coefficient positively correlate with rewarded performance (Figure 3.1H). That is, the more Air mice used duration information from their recent experience, the better they performed at the task. This was not present in CIE mice (Figure 3.1I). Collectively these results suggest that CIE impaired the ability for mice to use recent experience to guide subsequent lever presses and drive efficient performance, a hallmark of flexible behavior.

Figure 3.1. Prior CIE impairs the use of recent experience to guide flexible behavior. **(A)** CIE (Chronic Intermittent Ethanol) vapor exposure timeline. Training began 4 days (d) post final vapor exposure. **(B)** Hold down task schematic, mice had to press and hold down the lever for at least a minimum duration to earn food reward, without trials or cues. **(C)** Sample hold down session data, showing lever pressing durations across a single session for individual AIR and CIE mice. **(D)** Mean number of Total Lever Presses across days for all mice (n = 8 AIR, n = 8 CIE). 800ms indicates days with a duration criterion of >800ms, while 1600ms indicates days with a criterion of >1600ms. There was a main effect of training day during both 800ms (2-way RM ANOVA, (Day x Group): $F_{5,70} = 6.94$, $p < 0.0001$) and 1600ms training days ($F_{5,70} = 9.45$, $p < 0.0001$) but no group differences nor an interaction. **(E)** The % of lever presses that met criteria across days. Main effects of Day during both the 800ms ($F_{5,70} = 8.12$, $p < 0.0001$) and 1600ms days ($F_{5,70} = 9.45$, $p < 0.0001$). **(F)** The distribution of lever press durations on the final (i.e., the 6th) 800ms and final 1600ms training days. A 3-way RM ANOVA (Duration Bin x Group x Criterion) revealed a main effect only of Duration Bin ($F_{15,420} = 74.6$, $p < 0.001$), and a Duration Bin x Criterion interaction ($F_{15,420} = 9.17$, $p < 0.001$). **(G)** A linear mixed effect model predicting the duration of the current lever press (n) given the durations of prior lever presses (n-back). Here we show Beta (β) coefficients for the individual n-back presses. A 2-way ANOVA (n-back x Group) revealed main effects of both n-back ($F_{9,500960} = 18.6$, $p < 0.0001$) and Group ($F_{1,500960} = 26.4$, $p < 0.0001$), and an interaction ($F_{9,500960} = 8.56$, $p < 0.0001$). Bonferroni-corrected post-hoc tests found a significant Group difference only for press n - 1 ($t_{500960} = 8.76$, $p < 0.0001$). **(H)** β coefficients for linear mixed effect models built using individual session data and correlated with the %Presses that Met Criteria in AIR mice. Shades show individual subject data across days. A repeated measures correlation revealed a significant relationship (RM r = 0.285, DF = 87, slope = 0.002, p = 0.0069). **(I)** As in (H), except for CIE mice, where there was no significant relationship between β and %Presses Met (RM r = -0.038, DF = 87, slope = -0.0003 p = 0.72). Data points represent Mean ± SEM. n.s. = Not Significant. ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

Figure 3.1

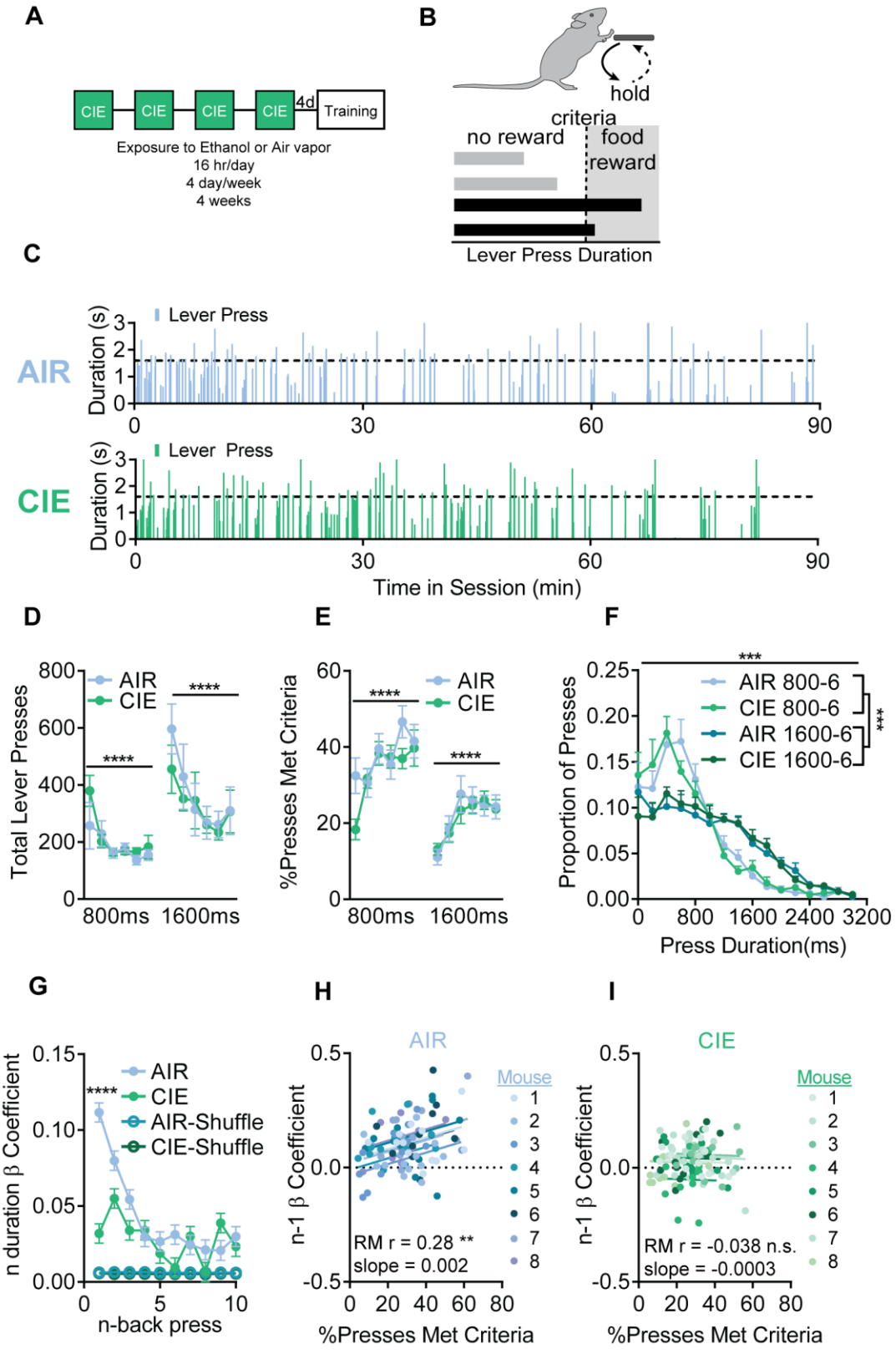

A
CIE
CIE
CIE
CIE
4d
Training
Exposure to Ethanol or Air vapor
16 hr/day
4 day/week
4 weeks
B
hold
criteria
no reward
food reward
Lever Press Duration
C
AIR
Lever Press
Duration (s)
CIE
Lever Press
Time in Session (min)
D
Total Lever Presses
AIR
CIE
800ms
1600ms
E
%Presses Met Criteria
F
Proportion of Presses
AIR 800-6
CIE 800-6
AIR 1600-6
CIE 1600-6
Press Duration(ms)
G
n duration β Coefficient
AIR-Shuffle
CIE-Shuffle
n-back press
H
AIR
n-1 β Coefficient
Mouse
RM r = 0.28 **
slope = 0.002
%Presses Met Criteria
I
CIE
RM r = -0.038 n.s.
slope = -0.0003

Figure 3.S1

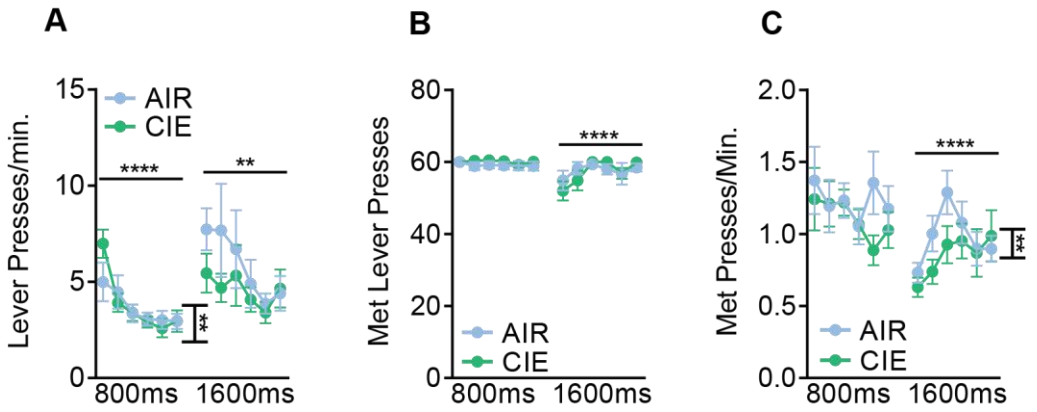


Figure 3.S1. No sustained difference in rate of lever pressing between Groups. **(A)** Mean rate of total lever presses (presses/minute) across days of training. During 800ms training, a 2-way RM ANOVA (Day x Group) revealed a main effect of day ($F_{5,70} = 21.0$, $p < 0.0001$), and a Day x Group interaction ($F_{5,70} = 3.07$, $p = 0.015$). No individual days differed in post hoc testing. Further, during 1600ms training, there was a main effect only of Day ($F_{5,70} = 4.03$, $p = 0.0028$). **(B)** Average number of Met Lever Presses (i.e., rewarded presses) across days. Main effect only of day, only during 1600ms training (2-way RM ANOVA, (Day x Group): $F_{5,70} = 5.92$, $p = 0.0001$). **(C)** Rate of Met lever presses (per minute). There were no main effects or interactions during 800ms training. During 1600ms training 2-way RM ANOVA (Day x Group) revealed a main effect of day ($F_{5,70} = 13.5$, $p < 0.0001$) and an interaction ($F_{5,70} = 4.19$, $p = 0.0022$), though no individual days differed in post hoc testing. Collectively, we see only transient and inconsistent differences (increased rate in CIE in A, but decreased in C) in rate of responding between AIR and CIE mice. Data points are Mean ± SEM. 800ms = days with a >800ms criterion, 1600ms = days with a >1600ms criterion. * = $p < 0.05$, ** = $p < 0.01$, **** = $p < 0.0001$.

Prior CIE increases the excitability of M2 projection neurons

We next sought to investigate how prior CIE affected the intrinsic properties of M2 projection neurons, which have been implicated in flexible behavior (Morris, Kundu, et al., 2016; Schreiner & Gremel, 2018; Siniscalchi et al., 2016, 2019). We performed *ex vivo* slice electrophysiological recordings of M2 projection neurons 1-3 weeks post CIE. Prior CIE significantly increased the excitability of M2 projection neurons (Figures 3.2A-B). Although we did not aim to examine the time course of effects, there was no obvious relationship between time from the last CIE exposure and excitability levels (Figure 3.S2A). There was a significant difference in input resistance (Figure 3.S1B-D), and no differences in resting membrane potential, first spike latency, threshold, afterhyperpolarization, amplitude, rise time, or half width (Table 3.S1). Thus, CIE seems to cause long-lasting increases in the intrinsic activity of M2 projection neurons.

Figure 3.2

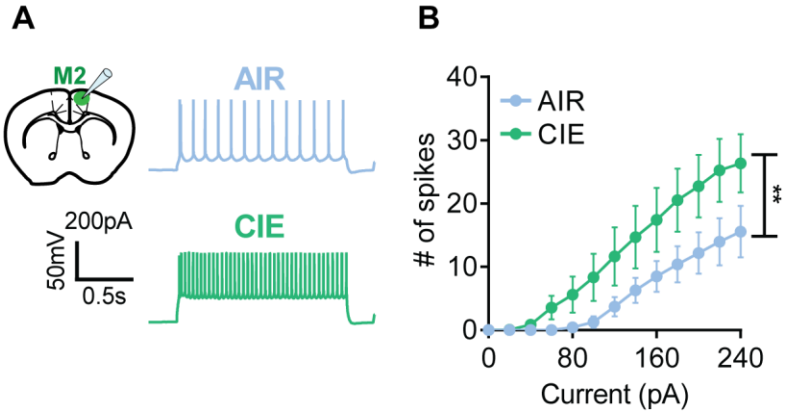


Figure 3.2. Prior CIE increases the excitability of M2 projection neurons. **(A)** (left) schematic of M2 *ex vivo* electrophysiological recording site and (right) representative traces of action potential firing at 200 picoamps (pA). **(B)** Average number (#) of spikes plotted against the amount of injected current. AIR n = 9 cells, 4 mice; CIE n = 10 cells, 7 mice. 2-way RM ANOVA (Group x Current), main effect of current ($F_{12,204}$ = 31.9, $p < 0.0001$) and an interaction ($F_{12,204}$ = 2.30, $p = 0.0091$). mV = millivolts. Data represent Mean ± SEM. ** = $p < 0.01$.

Figure 3.S2

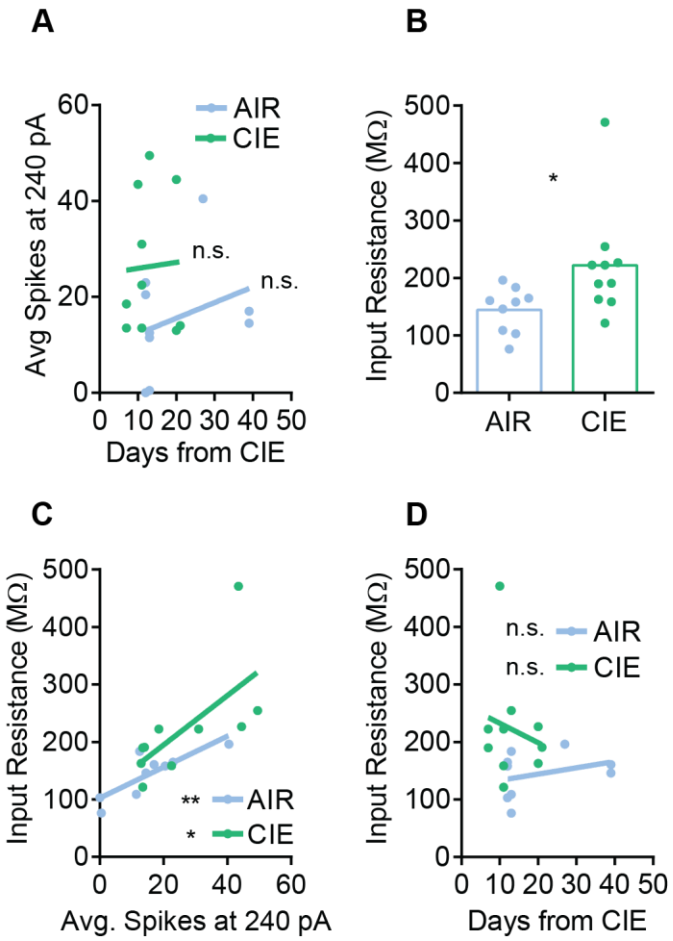


Figure 3.S2. Vapor exposure induces long-lasting changes in M2 excitability and input resistance. **(A)** Average spikes at 240 pA (picoamps) plotted against days from final vapor exposure. There was no significant linear relationship in either AIR (slope = 0.33, F = 0.77, p = 0.41), or CIE (slope = 0.12 $F_{1,8}$ = 0.02, p = 0.90) cells. **(B)** Input resistance in Megaohms (MΩ). There was a significant increase in input resistance in CIE cells (unpaired t-test with Welch's correction $t_{12.3}$ = 2.35, p = 0.036). **(C)** Input resistance plotted against excitability at 240 pA. There was a significant linear relationship in both AIR (slope = 2.69, F = 15.1, p = 0.006, R^2 = 0.68), and CIE (slope = 0.24, F = 5.97, p = 0.041, R^2 = 0.43), but no difference in the slopes (F = 0.55) or intercepts (F = 1.75) between AIR/CIE. **(D)** Input Resistance plotted against days from final vapor exposure. No significant relationship in either AIR (slope = 1.1, F = 0.81, p = 0.40), or CIE (slope = -3.37, F = 0.29, p = 0.61). n.s. = Not Significant. * = $p < 0.05$, ** = $p < 0.01$.

Prior CIE induces hyperactive calcium activity of M2-DMS projection neurons and uncouples activity and behavior

To address if and how CIE affects M2-DMS activity, modulation, and function during flexible behavior, we used a dual virus cre-dependent strategy to express the fluorescent calcium indicator GCaMP6s only in M2-DMS projection neurons, and implanted optical ferrules centered on M2 prior to CIE exposure (Figure 3.3A). We then recorded population calcium activity of M2-DMS projection neurons in Air and CIE mice using *in vivo* fiber photometry during task performance. Peak analysis of the session-long calcium signal (Pribiag et al., 2021) showed that prior CIE led to a significant increase in overall calcium transients (Figures 3.3B-C).

Prior CIE may also lead to alterations in the recruitment and/or modulation of calcium activity during task performance. We aligned calcium activity to the onset, duration, and offset of lever pressing. Figure 3.3D shows the varied modulation of calcium activity across lever pressing in a representative Air and CIE mouse that is significantly higher in the CIE-exposed mouse (assessed via running permutation tests, see Methods and (Jean-Richard-dit-Bressel et al., 2020)). We z-scored and averaged the calcium activity across all mice relative to a baseline period (-15s to -5s prior to press onset). Overall, CIE led to increased calcium activity throughout the entire press-aligned window (Figure 3.3E). We also alternatively analyzed this data using average traces per mouse (Figure 3.S3A), or calculating calcium activity as DF/F with a session-long mean to control for baseline periods which included prior lever presses (Figure 3.S3B). Both analyses yielded similar results.

The increased calcium activity in both Air and CIE mice just after press offset (Figure 3.3E, right) may reflect altered reward-related processing. Therefore, we segmented all lever presses based on whether they Met or Failed to meet the duration criteria. In both Air and CIE mice, the increased calcium activity after press offset was selective for Met presses (Figure 3.3F, right). There were also Met/Fail differences in calcium activity both before press onset and during action execution itself (Figure 3.3F). Thus, M2-DMS activity is differentially modulated by press durations before, during, and after those durations occur, suggesting M2-DMS activity may predict and encode press durations. Further, the average difference between Met/Fail lever presses was larger in CIE mice than Air mice at all three timepoints (Figure 3.3G). This raises the hypothesis that CIE alters the relationship between M2-DMS activity and press duration. To directly investigate this, we trained a linear SVM decoder to predict lever press duration using calcium activity data with 10 k-fold cross validations. We found significant decoding accuracy (chance at 25%) of lever press duration from M2-DMS activity, and this decoding accuracy significantly reduced in CIE relative to Air mice (Figure 3.3H). Collectively, these results suggest that CIE induced hyperactivity of M2-DMS projections neurons, which eroded the usual relationship between M2-DMS activity and behavioral output.

Figure 3.3. Prior CIE induces hyperactive calcium activity of M2-DMS projection neurons and uncouples activity and behavior. **(A)** (top) Schematic of dual virus targeting of GCaMP6s to M2-DMS projection neurons to n = 6 AIR and n = 8 CIE exposed mice, and (below) representative spread and fiber placement. **(B)** Representative calcium activity traces showing calcium transient analysis. Calcium signals above the dashed line were subjected to peak analysis, and those identified as events are indicated with tick marks above. **(C)** Average calcium events (in Hertz) across a session in AIR and CIE mice. There was a significantly increased rate of events in CIE mice (Mann-Whitney test, U = 449, AIR n = 46 (sessions), CIE n = 61 (sessions), $p < 0.0001$). **(D)** Representative calcium activity from sample mice showing calcium activity z-scored to a baseline period (-15s to -5s prior to press onset) aligned to the onset (left) duration (center) and offset (right) of lever pressing. Black bars indicate timepoints for which there is a significant AIR/CIE difference (Sig.) determined via permutation tests. **(E)** As in (D) except showing data from all mice collapsed together (lever presses n = 11,200 AIR, n = 14,792 CIE). **(F)** As in (E) except segmenting out lever presses based on whether they met the criterion (Met) or failed to do so (Fail) in both AIR and CIE mice. Here, significance markers indicate permutation tests comparing Met versus Fail activity within AIR (AIR-Met/Fail Sig.) and CIE (CIE-Met/Fail Sig.) mice. **(G)** The mean difference (Met - Fail) in the Met/Fail traces from (F). AIR and CIE mice differed at all three event windows (Mann-Whitney tests, Onset: U = 447, AIR/CIE n = 61 (samples, collected at 20 Hz for 3s), $p < 0.0001$. Duration: U = 72, AIR/CIE n = 20 (samples), $p = 0.0003$. Offset: U = 1303, AIR/CIE n = 61 (samples), $p = 0.0041$). **(H)** Decoding accuracy using calcium activity to predict press duration quartile. Data points are accuracy from individual sessions, dotted line indicates chance performance (25%). AIR and CIE mice significantly differed in decoding accuracy (unpaired t-test, $t_{105} = 3.19$, $p = 0.0019$). Data in (D) and (G) represent Mean $\pm$ SEM, while data in (E-F) are Mean $\pm$ 99% bootstrapped confidence intervals. Sig. = Significant difference. ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

Figure 3.3

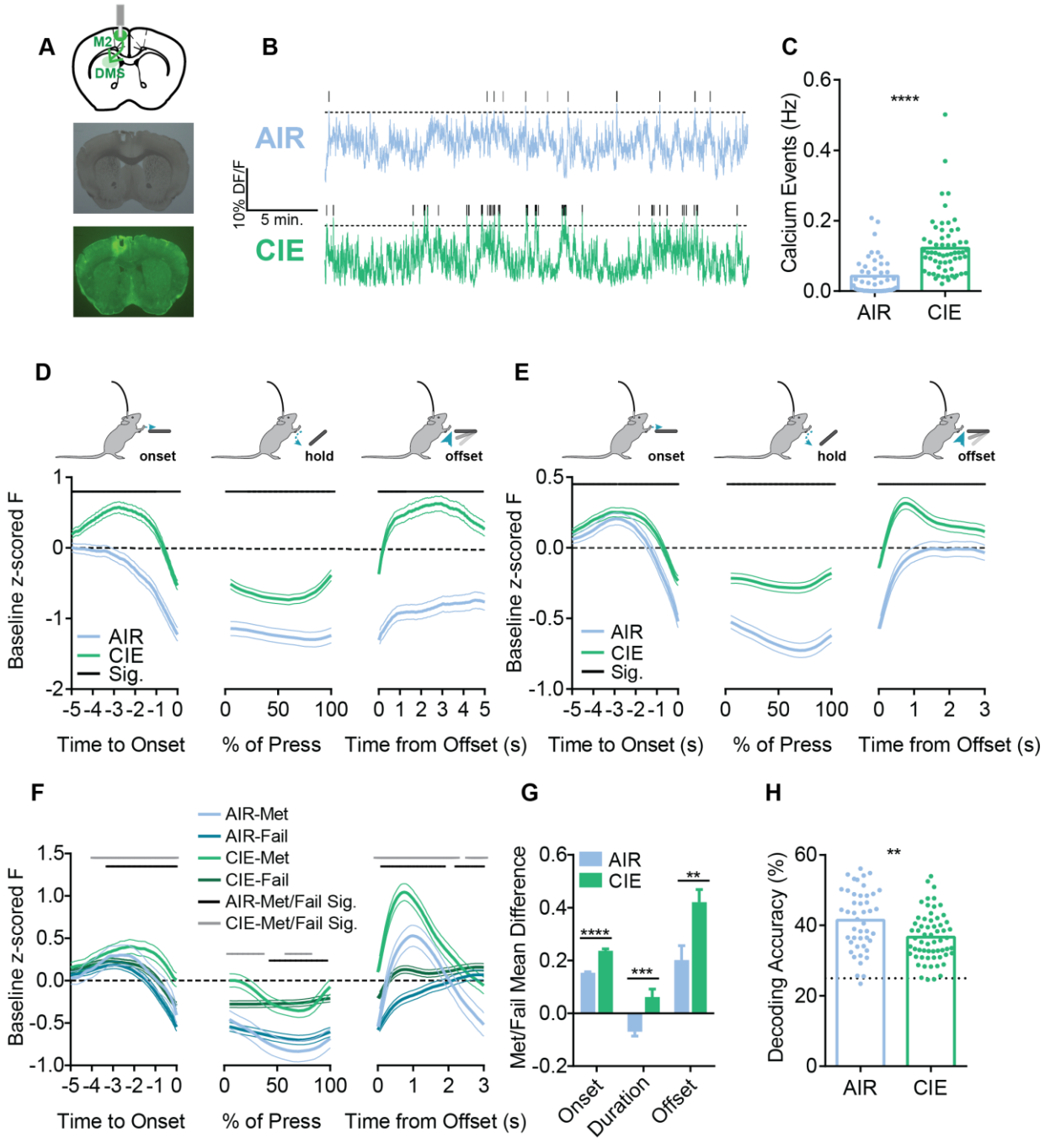

A
M2
DMS
B
AIR
10% DF/F
5 min.
CIE
C
Calcium Events (Hz)
0.6
0.4
0.2
0.0
****
AIR
CIE
D
onset
hold
offset
Baseline z-scored F
AIR
CIE
Sig.
Time to Onset
% of Press
Time from Offset (s)
E
onset
hold
offset
Baseline z-scored F
AIR
CIE
Sig.
Time to Onset (s)
% of Press
Time from Offset (s)
F
AIR-Met
AIR-Fail
CIE-Met
CIE-Fail
AIR-Met/Fail Sig.
CIE-Met/Fail Sig.
Baseline z-scored F
Time to Onset
% of Press
Time from Offset (s)
G
AIR
CIE
Met/Fail Mean Difference
****
***
**
Onset
Duration
Offset
H
Decoding Accuracy (%)
**
AIR
CIE

Figure 3.S3

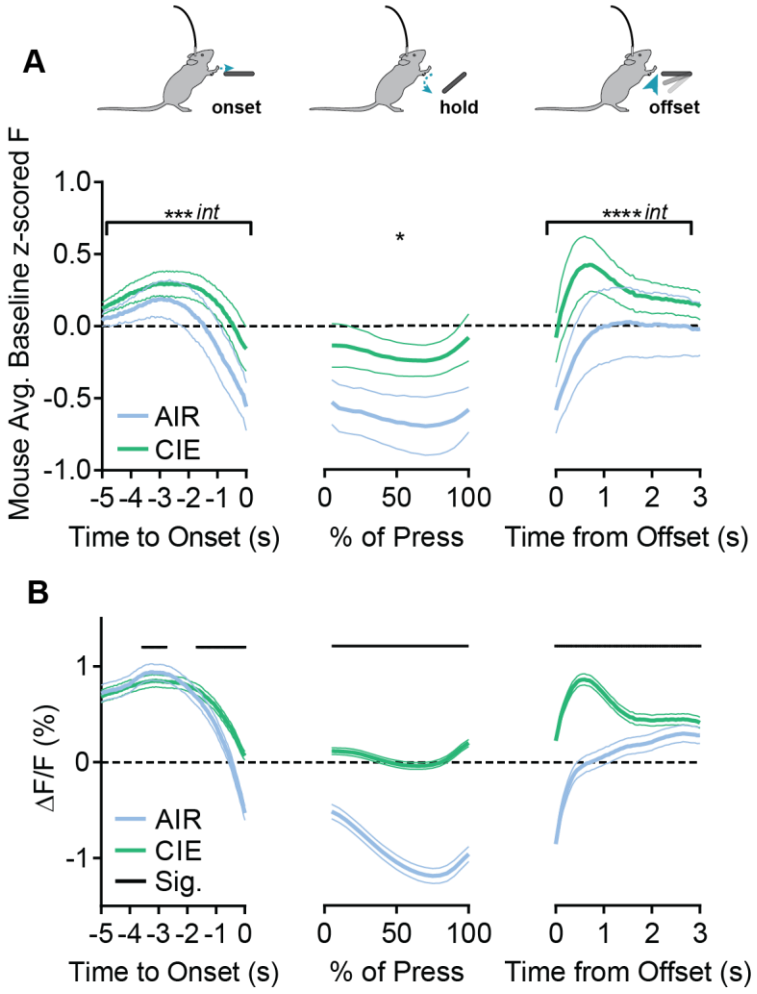


Figure 3.S3. Alternative calcium activity analyses still show hyperactive M2-DMS in CIE-exposed mice. **(A)** As in Figure 3E, calcium activity is aligned to press onset (left) duration (middle) and offset (right), only showing mean baseline z-scored values per individual mice (n = 6 AIR, n = 8 CIE), rather than collapsing across all mice (note intra-subject variance is lost in this measurement). 2-way RM ANOVAs (Group x Time point) reveal at onset: a main effect of Time point ($F_{100,1200} = 15.2$, $p < 0.0001$) and an interaction ($F_{100,1200} = 1.56$, $p = 0.0006$), during lever pressing: a main effect only of Group ($F_{19,228} = 1.41$, $p = 0.046$), and at offset: a main effect of Time point ($F_{60,720} = 4.07$, $p < 0.0001$) and an interaction ($F_{60,720} = 2.41$, $p < 0.0001$). **(B)** As in Figure 3E, only using ΔF/F with a session-long mean rather than z-scoring calcium activity relative to a baseline window. Thus, calcium activity still differs between AIR and CIE-exposed mice when using ΔF/F or mouse average measures. Black bars in (B) indicate significant Group differences assessed via running permutation test. Data are Mean ± SEM in (A), and Mean ± 99% Bootstrapped CI in (B). Sig. = Significant Difference. int = Significant Interaction. * = $p < 0.05$, *** = $p < 0.001$, **** = $p < 0.0001$.

Chemogenetic inhibition of hyperactive M2-DMS rescues behavioral flexibility

CIE disrupted M2-DMS *in vivo* activity and behavioral flexibility, but it was unclear whether M2-DMS activity changes causally led to the observed deficit in behavioral flexibility. To examine this, we applied a chemogenetic approach to rescue M2-DMS hyperactivity resulting from CIE. We used a dual viral vector strategy to target expression of the cre-dependent inhibitory chemogenetic receptor hM4Di (H4) or the cre-dependent fluorophore tdTomato as a control (Ctl) to M2-DMS projection neurons in both AIR and CIE exposed mice (Figure 3.4A, Figure 3.S4A), giving four groups of comparison (AIR Ctl, AIR H4, CIE Ctl, and CIE H4). All mice received the H4-agonist CNO 30 minutes prior to every lever hold down session (Figure 3.4B). In a subset of mice, we verified that CNO application reduced the excitability of M2-DMS neurons only in H4-expressing mice via *ex vivo* slice electrophysiology (Figure 3.S4B).

As in our other manipulations, there was little difference among groups in coarse measurements of behavior including Total Lever Presses (Figure 3.4C), %Presses Met Criteria (Figure 3.4D), and Lever Press Durations on the final training day (Figure 3.4E). The one exception was a significant Air/CIE group difference in %Presses Met Criteria during 1600ms training, where there was a significant Air/CIE difference, but not H4/Ctl difference, nor a significant interaction between these factors. Furthermore, there were no group differences in the rate of lever pressing, nor in the number or rate of met criteria lever presses (Figures 3.S4 C-E). Thus, neither CIE nor M2-DMS inhibition led to large changes in coarse behavior, but again - mice may reach largely similar levels of performance using different behavioral strategies.

We once more built LMEs to determine if mice were using their recent experience to guide behavior (Figure 3.4F). We found a significant interaction between Vapor Exposure and H4/Ctl group on the magnitude of the n - 1 β coefficient (no main effects). Post-hoc comparisons showed a replication of our initial finding, with Air Ctl mice having a significantly larger magnitude n - 1 β coefficient relative to CIE Ctl mice. Further, H4 expression in Air mice led to a significantly reduced magnitude n - 1 β coefficient relative to that observed in Air Ctl mice, showing M2-DMS activity contributes to the use of recent experience. In contrast, comparing CIE H4 mice to CIE Ctl mice (i.e., the treatment group), we found an *increase* in the magnitude of the n - 1 β coefficient in CIE H4 mice. Indeed, the n - 1 β coefficient magnitude in CIE H4 mice did not differ that of Air Ctl animals. Thus, by reducing CIE-induced M2-DMS hyperactivity, we rescued the use of recent experience (recently executed durations) to flexibly guide behavior.

Figure 3.4. Chemogenetic inhibition of hyperactive M2-DMS rescues behavioral flexibility. **(A)** Schematic of chemogenetic inhibition of M2-projection neurons (left) and representative viral spread (right). **(B)** Timeline of experiments. 1 week (w) after surgery, mice underwent 4 rounds of vapor exposure. 4 days (d) after exposure, mice began pretraining, and then were introduced to the hold down task with a criterion of >800ms for 6 days, followed by >1600ms for 6 days. Each day of hold down training, all mice were given intraperitoneal (i.p.) injections of the H4-agonist Clozapine-N-Oxide (CNO, 10 mg/ml) 30 minutes prior to the start of the session. **(C)** Mean total lever presses across days of training (AIR Ctl n = 6, AIR H4 n = 8, CIE Ctl n = 6, CIE H4 n = 7). Main effect only of day (3-way RM ANOVA, Vapor Group x DREADD Group x Day) during both 800ms training ($F_{5,115} = 73.4$, $p < 0.001$) and 1600ms days ($F_{5,115} = 22.8$, $p < 0.001$). **(D)** Average % of presses that met criteria across days of training. 3-way RM ANOVA (Vapor Group x DREADD Group x Day) showed only a main effect of day during 800ms training ($F_{5,115} = 99.2$, $p < 0.001$). During the 1600ms training there was a main effect of day ($F_{5,115} = 14.0$, $p < 0.001$), as well as Vapor Group ($F_{1,23} = 4.33$, $p = 0.049$), but there was no main effect of DREADD Group, nor an interaction between these two factors. **(E)** Distribution of lever press durations on the final day of 1600ms training, with a 3-way RM ANOVA (Vapor Group x DREADD Group x Duration Bin) showing a main effect only of Duration Bin ($F_{15,345} = 43.2$, $p < 0.001$). Dashed line indicates the 1600ms criterion. **(F)** Magnitude of the n - 1 Beta (β) coefficient from a linear mixed effect model using n-back durations to predict n press duration. A 2-way ANOVA (Vapor Group x DREADD Group) found no main effects, but did find a significant interaction ($F_{1,85311} = 31.2$, $p < 0.0001$). Bonferroni-corrected post hoc testing revealed that there were significant differences between: AIR Ctl and CIE Ctl ($t_{85311} = 3.85$, $p = 0.0007$), AIR Ctl and AIR H4 ($t_{85311} = 3.77$, $p = 0.0010$), CIE Ctl and CIE H4 ($t_{85311} = 4.14$, $p = 0.0002$), and between AIR H4 and CIE H4 ($t_{85311} = 4.13$, $p = 0.0002$). Data are Mean $\pm$ SEM. * = $p < 0.05$, *** = $p < 0.001$, **** = $p < 0.0001$.

Figure 3.4

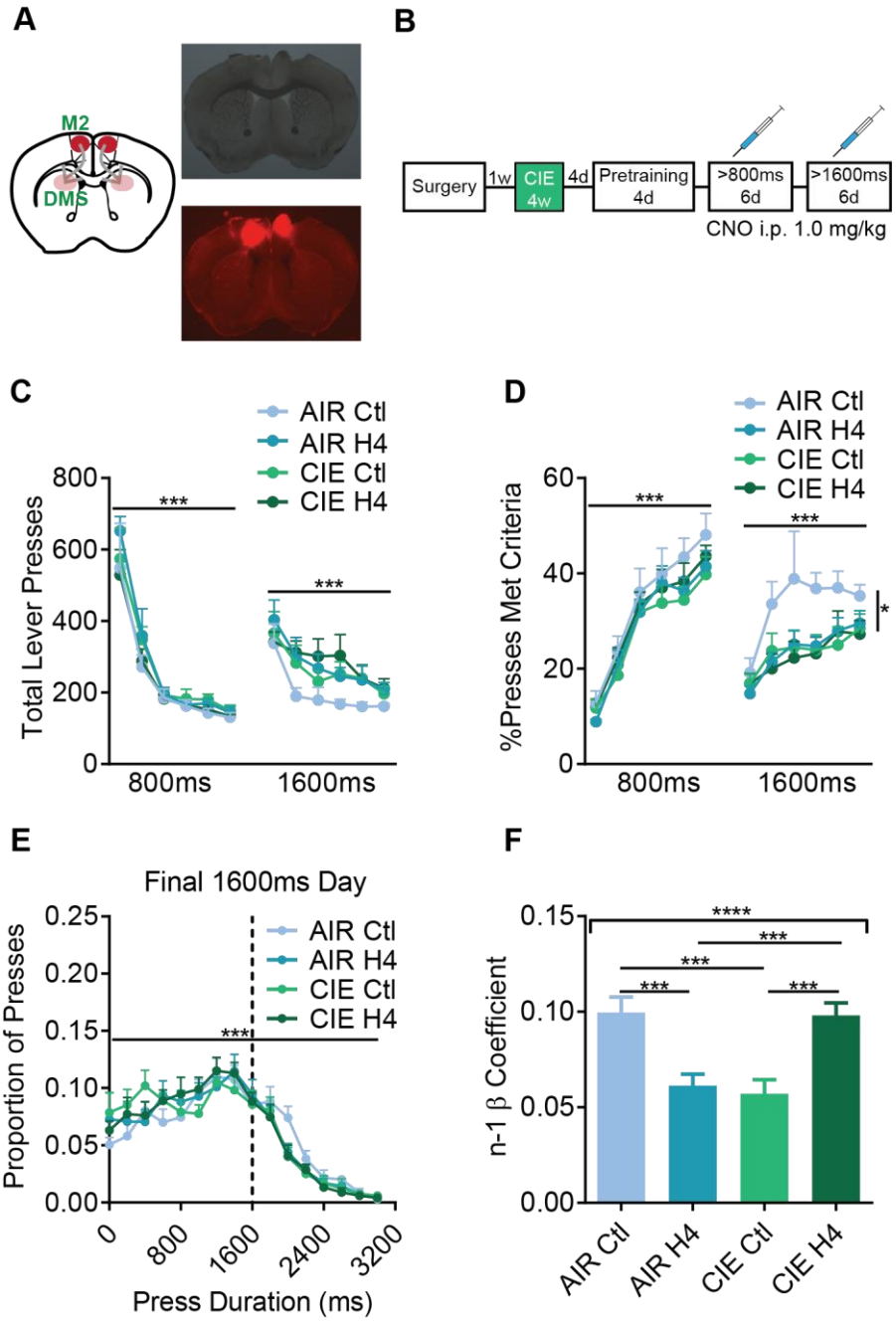

A
M2
DMS
B
Surgery
1w
CIE
4w
4d
Pretraining
4d
>800ms
6d
>1600ms
6d
CNO i.p. 1.0 mg/kg
C
AIR Ctl
AIR H4
CIE Ctl
CIE H4
Total Lever Presses
800ms
1600ms
D
%Presses Met Criteria
E
Final 1600ms Day
Proportion of Presses
Press Duration (ms)
F
n-1 β Coefficient

Figure 3.S4

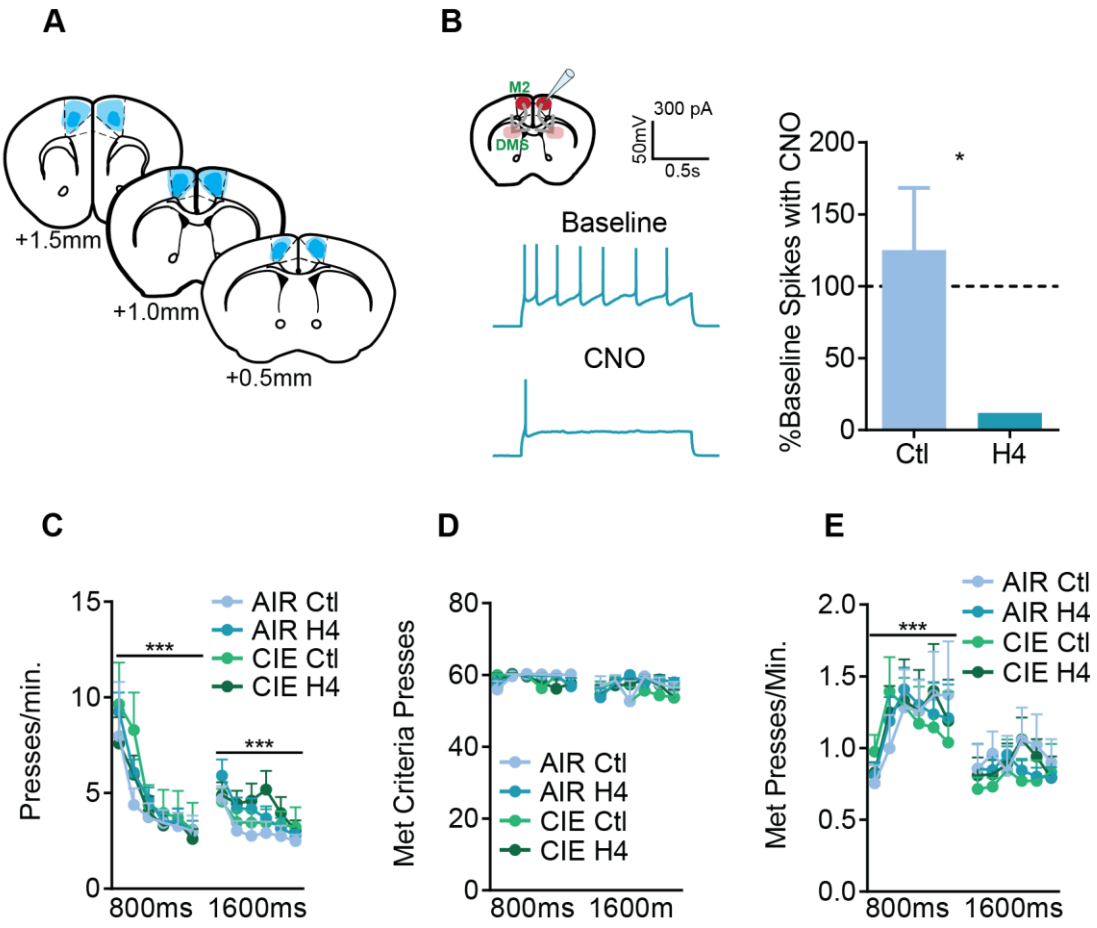


Figure 3.S4. Viral spread, DREADD verification, and rate of lever pressing for chemogenetic inhibition experiment. **(A)** Average (dark blue) and maximal (light blue) spread of H4 expression, assessed via mCherry fluorescent tag. Distances (+1.5, +1.0, and +0.5mm) are anterior relative to Bregma. **(B)** (left) sample trace of a cell expressing the inhibitory DREADD receptor (H4) at Baseline and after washing on the H4-agonist Clozapine-N-Oxide (CNO, 10uM). pA = Picoamps. mV = Milivolts. s = seconds. (right) %Change in spiking as a result of CNO relative to baseline at 300 pA in 1 Control (Ctl) cell and 1 H4 cell (cells are from 2 AIR exposed mice). Data represent 3 replicates. An unpaired t-test revealed a significant Group difference (t_4 = 4.52, p = 0.011). **(C)** Mean rate of total lever presses per minute. 3-way RM ANOVAs (Day x Vapor Group x DREADD Group) revealed a main effect only of Day during both 800ms ($F_{5,115}$ = 42.2, p < 0.001) and 1600ms training ($F_{5,115}$ = 10.4, p < 0.001). **(D)** Mean number of Met Criteria Presses across training. There were no main effects of Day, Vapor Group, or DREADD Group. **(E)** Mean rate of Met Criteria Presses (per minute) across training. 3-way RM ANOVA (Day x Vapor Group x DREADD Group)

revealed a main effect only of Day, only during 800ms training ($F_{5,115} = 8.41$, $p < 0.001$). Data are Mean ± SEM. * = $p < 0.05$, *** = $p < 0.001$.

Discussion

Premotor corticostriatal regions in humans and rodents are thought to support behavioral flexibility (Morris, Kundu, et al., 2016), and their disruption is linked to compulsive disorders (V. L. Corbit et al., 2019; Gomes et al., 2012; Hawken et al., 2016; Mantovani et al., 2013). While there have been reports of alterations to premotor circuits in AUD (Claus et al., 2011; Duka et al., 2011; Morris et al., 2018; Sjoerds et al., 2014), there has been a dearth of insight into the precise neural mechanisms and the behavioral consequences of such disruption. Here we show that prior chronic alcohol exposure reduces the use of recent experience to flexibly guide behavior due to an alcohol-induced increase in the activity of M2 and M2-DMS circuits. Thus our findings reveal one specific circuit through which chronic alcohol can disrupt behavioral flexibility and support the targeting of Pre-SMA/SMA to treat altered executive function in AUD.

Substantial evidence suggests that AUD can disrupt behavioral flexibility (Claus et al., 2011; Duka et al., 2011; Scaife & Duka, 2009; Shnitko et al., 2020; Sjoerds et al., 2014) and goal-directed control (Barker et al., 2015; L. H. Corbit & Janak, 2016; Dickinson et al., 2002; Everitt & Robbins, 2005; Renteria et al., 2018; Sjoerds et al., 2013). However, a host of different computational mechanisms support flexibility and goal-directed decision-making (Schreiner et al., 2020; Shnitko et al., 2020). By using a continuous task, we were able to show that one specific aspect of flexibility - the use of recent experience to guide behavior - was disrupted by AUD and controlled by M2-

DMS projection neurons. Our results have important implications for AUD; we saw a disruption in experience-based computations for a non-drug reward in protracted withdrawal. This suggests long-lasting neuroadaptations in M2 arising in response to chronic alcohol exposure and withdrawal, as evidenced by increased calcium activity, excitability, and input resistance. These long-lasting changes may contribute to deficits in behavioral flexibility observed in AUD, including impulsivity and impaired response inhibition (Claus et al., 2011; Duka et al., 2011; Sjoerds et al., 2014). Impaired sensitivity to recent experience was causally tied to M2 and M2-DMS activity. This supports and extends previous correlative evidence in humans showing AUD is associated with disruption to both premotor regions and flexibility (Claus et al., 2011; Duka et al., 2011; Sjoerds et al., 2014). We report increases in M2-DMS activity, while prior studies of premotor and prefrontal cortex function in humans with AUD variously report increases and decreases in activity (Claus et al., 2011; Duka et al., 2011; Morris et al., 2018; Sjoerds et al., 2014). This highlights the need to examine activity in a nuanced manner in relation to the computations being performed in order to understand the functional consequences. This is likely to be especially important for precise targeting of novel region-specific treatments.

To our knowledge, this is the first study investigating the function of M2 neurons that project to dorsal striatum in the context of chronic alcohol. Hypotheses related to habitual control in AUD (Gremel & Lovinger, 2017; Lovinger & Gremel, 2021) have suggested increased motor and sensory input into the dorsal striatum may contribute to habit-related phenotypes. The present findings provide some support for this; chronic alcohol induced hyperactivity of M2-DMS calcium activity *in vivo* during task

performance and disrupted behavioral flexibility. However, we found weaker decoding of press durations from this hyperactivity and chemogenetically inhibiting this hyperactivity restored the usual activity-duration relationship. One of the most intriguing findings of the present study was that both increases (from prior CIE) and decreases (from chemogenetic inhibition) in M2-DMS activity led to similar behavioral alterations. While the present work did not examine whether *in vivo* hyperactivity translated into increased M2-DMS transmission, this does suggest that the potential increased drive from M2-DMS following CIE does not *directly* support the observed phenotype, since reduced drive also leads to a similar phenotype. Rather, there may be an optimal level, amount, or timing of M2-DMS activity, such that any shift may erode the fidelity of the activity-duration relationship. This adds to a growing body of literature suggesting that the patterning of neural activity is decisive and that there is not a simple linear relationship between brain activity and behavioral output (Tecuapetla et al., 2016).

How might prior chronic alcohol produce hyperactive M2-DMS? In part, this may reflect alterations within M2 itself, as we found that chronic alcohol altered the intrinsic properties of M2 projection neurons. Chronic alcohol may also affect local circuitry transmission and/or input into M2. One interesting candidate is the orbitofrontal cortex, as it is affected by alcohol (Moorman, 2018; Shields & Gremel, 2020) and its projection to M2 is implicated in behavioral flexibility (Johnson et al., 2016; Schreiner & Gremel, 2018). A further question is whether there are synapse-specific alterations of M2 projections onto dorsal striatum neurons (Renteria et al., 2018, 2021; Rothwell et al., 2015). Rodent models have implicated these M2-DMS

projections in mediating compulsivity in OCD (V. L. Corbit et al., 2019), as well as motor and working memory deficits in Parkinson's disease (Magno et al., 2019). Given the above findings, how chronic alcohol alters M2 recruitment of dorsal striatal circuitry for behavioral control remains to be investigated.

The human homologues of M2 - Pre-SMA/SMA - are dorsally located and accessible to region-specific treatments such as rTMS. Indeed, such treatments have shown promise in reducing compulsivity and improving behavioral flexibility in Obsessive Compulsive Disorder (Gomes et al., 2012; Hawken et al., 2016; Mantovani et al., 2013). These prior works demonstrate a preclinical to clinical translation for premotor circuits involved in psychiatric disease, and suggest that Pre-SMA/SMA may be fruitful therapeutic targets for the treatment of AUD in human patients. Here, we provide preclinical evidence supporting this potential treatment avenue, as well as mechanistic insight into the involved behavioral and neural controllers of behavioral flexibility that are disrupted by AUD.

Disclosures

The authors report no biomedical financial interests or potential conflicts of interest.

Acknowledgements

Chapter 3 is, in full, a reprint of the material to be submitted to *Biological Psychiatry*. Schreiner, Drew C.; Renteria, Rafael; Baltz, Emily T.; Cazares, Christian; Gremel, Christina M. The dissertation author was the primary investigator and author

of this paper. This project was funded by the NIH: F31AA027439 (D.C.S.), F32AA026776 (R.R.), R01AA026077 (C.M.G.), the NSF: DGE-1650112 (C.C.), DGE-2038238 (E.T.B), and a Whitehall Foundation Award (C.M.G).

References

Aron, A. R. (2011). From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. *Biological Psychiatry*, *69*(12), e55–e68. https://doi.org/10.1016/j.biopsych.2010.07.024

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a Cognitive Control Network Using Diffusion-Weighted Magnetic Resonance Imaging (MRI) and Functional MRI. *The Journal of Neuroscience*, *27*(14), 3743–3752. https://doi.org/10.1523/JNEUROSCI.0519-07.2007

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00456

Barker, J. M., Corbit, L. H., Robinson, D. L., Gremel, C. M., Gonzales, R. A., & Chandler, L. J. (2015). Corticostriatal circuitry and habitual ethanol seeking. *Alcohol*, *49*(8), 817–824. https://doi.org/10.1016/j.alcohol.2015.03.003

Barthas, F., & Kwan, A. C. (2017). Secondary Motor Cortex: Where 'Sensory' Meets 'Motor' in the Rodent Frontal Cortex. *Trends in Neurosciences*, *40*(3), 181–193. https://doi.org/10.1016/j.tins.2016.11.006

Becker, H. C., & Hale, R. L. (1993). Repeated Episodes of Ethanol Withdrawal Potentiate the Severity of Subsequent Withdrawal Seizures: An Animal Model of Alcohol Withdrawal "Kindling." *Alcoholism: Clinical and Experimental Research*, *17*(1), 94–98. https://doi.org/10.1111/j.1530-0277.1993.tb00731.x

Becker, H. C., & Lopez, M. F. (2004). Increased Ethanol Drinking After Repeated Chronic Ethanol Exposure and Withdrawal Experience in C57BL/6 Mice. *Alcoholism: Clinical and Experimental Research*, *28*(12), 1829–1838. https://doi.org/10.1097/01.ALC.0000149977.95306.3A

Becker, H. C., & Lopez, M. F. (2006). Increased Ethanol Drinking After Repeated Chronic Ethanol Exposure and Withdrawal Experience in C57BL/6 Mice. *Alcoholism: Clinical and Experimental Research*, *28*(12), 1829–1838. https://doi.org/10.1097/01.ALC.0000149977.95306.3A

Belin, D., Belin-Rauscent, A., Murray, J. E., & Everitt, B. J. (2013). Addiction: Failure of control over maladaptive incentive habits. *Current Opinion in Neurobiology*, *23*(4), 564–572. https://doi.org/10.1016/j.conb.2013.01.025

Carlson, V. C. C., Seabold, G. K., Helms, C. M., Garg, N., Odagiri, M., Rau, A. R., Daunais, J., Alvarez, V. A., Lovinger, D. M., & Grant, K. A. (2011). Synaptic and morphological neuroadaptations in the putamen associated with long-term, relapsing alcohol drinking in primates. *Neuropsychopharmacology*, *36*(12), 2513–2528.

Cazares, C., Schreiner, D. C., & Gremel, C. M. (2021). Different Effects of Alcohol Exposure on Action and Outcome Related Orbitofrontal Cortex Activity. *ENeuro*, *8*(2). https://doi.org/10.1523/ENEURO.0052-21.2021

Claus, E. D., Kiehl, K. A., & Hutchison, K. E. (2011). Neural and Behavioral Mechanisms of Impulsive Choice in Alcohol Use Disorder. *Alcoholism, Clinical and Experimental Research*, *35*(7), 1209–1219. https://doi.org/10.1111/j.1530-0277.2011.01455.x

Corbit, L. H., & Janak, P. H. (2016). Habitual Alcohol Seeking: Neural Bases and Possible Relations to Alcohol Use Disorders. *Alcoholism: Clinical and Experimental Research*, *40*(7), 1380–1389. https://doi.org/10.1111/acer.13094

Cuzon Carlson, V. C. (2018). GABA and Glutamate Synaptic Coadaptations to Chronic Ethanol in the Striatum. In K. A. Grant & D. M. Lovinger (Eds.), *The Neuropharmacology of Alcohol* (pp. 79–112). Springer International Publishing. https://doi.org/10.1007/164_2018_98

Corbit, V. L., Manning, E. E., Gittis, A. H., & Ahmari, S. E. (2019). Strengthened Inputs from Secondary Motor Cortex to Striatum in a Mouse Model of Compulsive Behavior. *Journal of Neuroscience*, *39*(15), 2965–2975. https://doi.org/10.1523/JNEUROSCI.1728-18.2018

Delevich, K., Okada, N. J., Rahane, A., Zhang, Z., Hall, C. D., & Wilbrecht, L. (2020). Sex and Pubertal Status Influence Dendritic Spine Density on Frontal Corticostriatal Projection Neurons in Mice. *Cerebral Cortex*, *30*(6), 3543–3557. https://doi.org/10.1093/cercor/bhz325

den Hartog, C., Zamudio-Bulcock, P., Nimitvilai, S., Gilstrap, M., Eaton, B., Fedarovich, H., Motts, A., & Woodward, J. J. (2016). Inactivation of the lateral orbitofrontal cortex increases drinking in ethanol-dependent but not non-dependent mice. *Neuropharmacology*, *107*, 451–459. https://doi.org/10.1016/j.neuropharm.2016.03.031

Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., Costa, R. M., & Sousa, N. (2009). Chronic Stress Causes Frontostriatal Reorganization and Affects Decision-Making. *Science*, *325*(5940), 621–625. https://doi.org/10.1126/science.1171203

Dickinson, A., Wood, N., & Smith, J. W. (2002). Alcohol seeking by rats: Action or habit? *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *55*(4), 331–348. https://doi.org/10.1080/0272499024400016

Dudek, M., Abo-Ramadan, U., Hermann, D., Brown, M., Canals, S., Sommer, W. H., & Hyytiä, P. (2015). Brain activation induced by voluntary alcohol and saccharin drinking in rats assessed with manganese-enhanced magnetic resonance

imaging. *Addiction Biology*, *20*(6), 1012–1021. https://doi.org/10.1111/adb.12179

Duka, T., Trick, L., Nikolaou, K., Gray, M. A., Kempton, M. J., Williams, H., Williams, S. C. R., Critchley, H. D., & Stephens, D. N. (2011). Unique Brain Areas Associated with Abstinence Control Are Damaged in Multiply Detoxified Alcoholics. *Biological Psychiatry*, *70*(6), 545–552. https://doi.org/10.1016/j.biopsych.2011.04.006

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, *8*(11), 1481–1489. https://doi.org/10.1038/nn1579

Fan, D., Rossi, M. A., & Yin, H. H. (2012). Mechanisms of Action Selection and Timing in Substantia Nigra Neurons. *Journal of Neuroscience*, *32*(16), 5534–5548. https://doi.org/10.1523/JNEUROSCI.5924-11.2012

Gomes, P. V. O., Brasil-Neto, J. P., Allam, N., & Rodrigues de Souza, E. (2012). A Randomized, Double-Blind Trial of Repetitive Transcranial Magnetic Stimulation in Obsessive-Compulsive Disorder With Three-Month Follow-Up. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *24*(4), 437–443. https://doi.org/10.1176/appi.neuropsych.11100242

Gomez, J. L., Bonaventura, J., Lesniak, W., Mathews, W. B., Sysa-Shah, P., Rodriguez, L. A., Ellis, R. J., Richie, C. T., Harvey, B. K., Dannals, R. F., Pomper, M. G., Bonci, A., & Michaelides, M. (2017). Chemogenetics revealed: DREADD occupancy and activation via converted clozapine. *Science*, *357*(6350), 503–507. https://doi.org/10.1126/science.aan2475

Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, *4*. https://doi.org/10.1038/ncomms3264

Gremel, C. M., & Lovinger, D. M. (2017). Associative and sensorimotor cortico-basal ganglia circuit roles in effects of abused drugs. *Genes, Brain and Behavior*, *16*(1), 71–85. https://doi.org/10.1111/gbb.12309

Hattori, R., Danskin, B., Babic, Z., Mlynaryk, N., & Komiyama, T. (2019). Area-Specificity and Plasticity of History-Dependent Value Coding During Learning. *Cell*, *177*(7), 1858-1872.e15. https://doi.org/10.1016/j.cell.2019.04.027

Hawken, E. R., Dilkov, D., Kaludiev, E., Simek, S., Zhang, F., & Milev, R. (2016). Transcranial Magnetic Stimulation of the Supplementary Motor Area in the Treatment of Obsessive-Compulsive Disorder: A Multi-Site Study. *International Journal of Molecular Sciences*, *17*(3). https://doi.org/10.3390/ijms17030420

Heilig, M., Egli, M., Crabbe, J. C., & Becker, H. C. (2010). Acute withdrawal, protracted

abstinence and negative affect in alcoholism: Are they linked? *Addiction Biology*, *15*(2), 169–184. https://doi.org/10.1111/j.1369-1600.2009.00194.x

Hintiryan, H., Foster, N. N., Bowman, I., Bay, M., Song, M. Y., Gou, L., Yamashita, S., Bienkowski, M. S., Zingg, B., Zhu, M., Yang, X. W., Shih, J. C., Toga, A. W., & Dong, H.-W. (2016). The mouse cortico-striatal projectome. *Nature Neuroscience*, *19*(8), 1100–1114. https://doi.org/10.1038/nn.4332

Hogarth, L. (2020). Addiction is driven by excessive goal-directed drug choice under negative affect: Translational critique of habit and compulsion theory. *Neuropsychopharmacology*, *45*(5), 720–735. https://doi.org/10.1038/s41386-020-0600-8

Jean-Richard-dit-Bressel, P., Clifford, C. W. G., & McNally, G. P. (2020). Analyzing Event-Related Transients: Confidence Intervals, Permutation Tests, and Consecutive Thresholds. *Frontiers in Molecular Neuroscience*, *13*. https://doi.org/10.3389/fnmol.2020.00014

Johnson, C. M., Peckler, H., Tai, L.-H., & Wilbrecht, L. (2016). Rule learning enhances structural plasticity of long-range axons in frontal cortex. *Nature Communications*, *7*, 10785. https://doi.org/10.1038/ncomms10785

Liu, W., & Crews, F. T. (2015). Adolescent Intermittent ethanol exposure enhances ethanol activation of the nucleus accumbens while blunting the prefrontal cortex responses in adult rat. *Neuroscience*, *293*, 92–108. https://doi.org/10.1016/j.neuroscience.2015.02.014

Lopes, G., Bonacchi, N., Frazão, J., Neto, J. P., Atallah, B. V., Soares, S., Moreira, L., Matias, S., Itskov, P. M., Correia, P. A., Medina, R. E., Calcaterra, L., Dreosti, E., Paton, J. J., & Kampff, A. R. (2015). Bonsai: An event-based framework for processing and controlling data streams. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00007

Lopez, M. F., & Becker, H. C. (2005). Effect of pattern and number of chronic ethanol exposures on subsequent voluntary ethanol intake in C57BL/6J mice. *Psychopharmacology*, *181*(4), 688–696. https://doi.org/10.1007/s00213-005-0026-3

Lovinger, D. M., & Alvarez, V. A. (2017). Alcohol and basal ganglia circuitry: Animal models. *Neuropharmacology*, *122*, 46–55. https://doi.org/10.1016/j.neuropharm.2017.03.023

Lovinger, D. M., & Gremel, C. M. (2021). A Circuit-Based Information Approach to Substance Abuse Research. *Trends in Neurosciences*, *44*(2), 122–135. https://doi.org/10.1016/j.tins.2020.10.005

Ma, T., Cheng, Y., Roltsch Hellard, E., Wang, X., Lu, J., Gao, X., Huang, C. C. Y.,

Wei, X.-Y., Ji, J.-Y., & Wang, J. (2018). Bidirectional and long-lasting control of alcohol-seeking behavior by corticostriatal LTP and LTD. *Nature Neuroscience*, *21*(3), 373–383. https://doi.org/10.1038/s41593-018-0081-9

Magno, L. A. V., Tenza-Ferrer, H., Collodetti, M., Aguiar, M. F. G., Rodrigues, A. P. C., Silva, R. S. da, Silva, J. do P., Nicolau, N. F., Rosa, D. V. F., Birbrair, A., Miranda, D. M., & Romano-Silva, M. A. (2019). Optogenetic stimulation of the M2 cortex reverts motor dysfunction in a mouse model of Parkinson's Disease. *Journal of Neuroscience*, *39*(17), 3234–3248. https://doi.org/10.1523/JNEUROSCI.2277-18.2019

Mantovani, A., Rossi, S., Bassi, B. D., Simpson, H. B., Fallon, B. A., & Lisanby, S. H. (2013). Modulation of motor cortex excitability in obsessive-compulsive disorder: An exploratory study on the relations of neurophysiology measures with clinical outcome. *Psychiatry Research*, *210*(3), 1026–1032. https://doi.org/10.1016/j.psychres.2013.08.054

Markowitz, J. E., Gillis, W. F., Beron, C. C., Neufeld, S. Q., Robertson, K., Bhagat, N. D., Peterson, R. E., Peterson, E., Hyun, M., Linderman, S. W., Sabatini, B. L., & Datta, S. R. (2018). The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell*, *174*(1), 44-58.e17. https://doi.org/10.1016/j.cell.2018.04.019

Moorman, D. E. (2018). The role of the orbitofrontal cortex in alcohol use, abuse, and dependence. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *87*, 85–107. https://doi.org/10.1016/j.pnpbp.2018.01.010

Morningstar, M. D., Linsenbardt, D. N., & Lapish, C. C. (2020). Ethanol Alters Variability, But Not Rate, of Firing in Medial Prefrontal Cortex Neurons of Awake-Behaving Rats. *Alcoholism: Clinical and Experimental Research*, *44*(11), 2225–2238. https://doi.org/10.1111/acer.14463

Morris, L. S., Baek, K., Kundu, P., Harrison, N. A., Frank, M. J., & Voon, V. (2016). Biases in the Explore–Exploit Tradeoff in Addictions: The Role of Avoidance of Uncertainty. *Neuropsychopharmacology*, *41*(4), 940–948. https://doi.org/10.1038/npp.2015.208

Morris, L. S., Baek, K., Tait, R., Elliott, R., Ersche, K. D., Flechais, R., McGonigle, J., Murphy, A., Nestor, L. J., Orban, C., Passetti, F., Paterson, L. M., Rabiner, I., Reed, L., Smith, D., Suckling, J., Taylor, E. M., Bullmore, E. T., Lingford-Hughes, A. R., Deakin, B., Nutt, D. J., Sahakian, B. J., Robbins, T. W., Voon, V., & ICCAM Consortium. (2018). Naltrexone ameliorates functional network abnormalities in alcohol-dependent individuals. *Addiction Biology*, *23*(1), 425–436. https://doi.org/10.1111/adb.12503

Morris, L. S., Kundu, P., Dowell, N., Mechelmans, D. J., Favre, P., Irvine, M. A., Robbins, T. W., Daw, N., Bullmore, E. T., Harrison, N. A., & Voon, V. (2016).

Fronto-striatal organization: Defining functional and microstructural substrates of behavioural flexibility. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *74*, 118–133. https://doi.org/10.1016/j.cortex.2015.11.004

Muñoz, B., Fritz, B. M., Yin, F., & Atwood, B. K. (2018). Alcohol exposure disrupts mu opioid receptor-mediated long-term depression at insular cortex inputs to dorsolateral striatum. *Nature Communications*, *9*(1), 1318. https://doi.org/10.1038/s41467-018-03683-1

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

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

Nimitvilai, S., Lopez, M. F., Mulholland, P. J., & Woodward, J. J. (2016). Chronic Intermittent Ethanol Exposure Enhances the Excitability and Synaptic Plasticity of Lateral Orbitofrontal Cortex Neurons and Induces a Tolerance to the Acute Inhibitory Actions of Ethanol. *Neuropsychopharmacology*, *41*(4), 1112–1127. https://doi.org/10.1038/npp.2015.250

Nimitvilai, S., Uys, J. D., Woodward, J. J., Randall, P. K., Ball, L. E., Williams, R. W., Jones, B. C., Lu, L., Grant, K. A., & Mulholland, P. J. (2017). Orbitofrontal Neuroadaptations and Cross-Species Synaptic Biomarkers in Heavy-Drinking Macaques. *Journal of Neuroscience*, *37*(13), 3646–3660. https://doi.org/10.1523/JNEUROSCI.0133-17.2017

Patton, M. S., Heckman, M., Kim, C., Mu, C., & Mathur, B. N. (2021). Compulsive alcohol consumption is regulated by dorsal striatum fast-spiking interneurons. *Neuropsychopharmacology*, *46*(2), 351–359. https://doi.org/10.1038/s41386-020-0766-0

Pereira, E. F., Aracava, Y., Aronstam, R. S., Barreiro, E. J., & Albuquerque, E. X. (1992). Pyrazole, an alcohol dehydrogenase inhibitor, has dual effects on N-methyl-D-aspartate receptors of hippocampal pyramidal cells: Agonist and noncompetitive antagonist. *Journal of Pharmacology and Experimental Therapeutics*, *261*(1), 331–340.

Platt, J. R., Kuch, D. O., & Bitgood, S. C. (1973). Rats' lever-press durations as psychophysical judgments of time. *Journal of the Experimental Analysis of Behavior*, *19*(2), 239–250. https://doi.org/10.1901/jeab.1973.19-239

Pribiag, H., Shin, S., Wang, E. H.-J., Sun, F., Datta, P., Okamoto, A., Guss, H., Jain,

A., Wang, X.-Y., De Freitas, B., Honma, P., Pate, S., Lilascharoen, V., Li, Y., & Lim, B. K. (2021). Ventral pallidum DRD3 potentiates a pallido-habenular circuit driving accumbal dopamine release and cocaine seeking. *Neuron*, *109*(13), 2165-2182.e10. https://doi.org/10.1016/j.neuron.2021.05.002

Renteria, R., Baltz, E. T., & Gremel, C. M. (2018). Chronic alcohol exposure disrupts top-down control over basal ganglia action selection to produce habits. *Nature Communications*, *9*(1), 211. https://doi.org/10.1038/s41467-017-02615-9

Renteria, R., Cazares, C., Baltz, E. T., Schreiner, D. C., Yalcinbas, E. A., Steinkellner, T., Hnasko, T. S., & Gremel, C. M. (2021). Mechanism for differential recruitment of orbitostriatal transmission during actions and outcomes following chronic alcohol exposure. *ELife*, *10*, e67065. https://doi.org/10.7554/eLife.67065

Rothermel, M., Brunert, D., Zabawa, C., Díaz-Quesada, M., & Wachowiak, M. (2013). Transgene Expression in Target-Defined Neuron Populations Mediated by Retrograde Infection with Adeno-Associated Viral Vectors. *Journal of Neuroscience*, *33*(38), 15195–15206. https://doi.org/10.1523/JNEUROSCI.1618-13.2013

Rothwell, P. E., Hayton, S. J., Sun, G. L., Fuccillo, M. V., Lim, B. K., & Malenka, R. C. (2015). Input- and Output-Specific Regulation of Serial Order Performance by Corticostriatal Circuits. *Neuron*, *88*(2), 345–356. https://doi.org/10.1016/j.neuron.2015.09.035

Scaife, J. C., & Duka, T. (2009). Behavioural measures of frontal lobe function in a population of young social drinkers with binge drinking pattern. *Pharmacology, Biochemistry, and Behavior*, *93*(3), 354–362. https://doi.org/10.1016/j.pbb.2009.05.015

Schreiner, D. C., & Gremel, C. M. (2018). Orbital Frontal Cortex Projections to Secondary Motor Cortex Mediate Exploitation of Learned Rules. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-29285-x

Schreiner, D. C., Renteria, R., & Gremel, C. M. (2020). Fractionating the all-or-nothing definition of goal-directed and habitual decision-making. *Journal of Neuroscience Research*, *98*(6), 998–1006. https://doi.org/10.1002/jnr.24545

Schreiner, D. C., Yalcinbas, E. A., & Gremel, C. M. (2021). A push for examining subjective experience in value-based decision-making. *Current Opinion in Behavioral Sciences*, *41*, 45–49. https://doi.org/10.1016/j.cobeha.2021.03.020

Shields, C. N., & Gremel, C. M. (2020). Review of Orbitofrontal Cortex in Alcohol Dependence: A Disrupted Cognitive Map? *Alcoholism: Clinical and Experimental Research*, *44*(10), 1952–1964. https://doi.org/10.1111/acer.14441

Shnitko, T. A., Gonzales, S. W., Newman, N., & Grant, K. A. (2020). Behavioral Flexibility in alcohol drinking monkeys: The morning after. *Alcoholism, Clinical and Experimental Research*, *44*(3), 729–737. https://doi.org/10.1111/acer.14289

Siniscalchi, M. J., Phoumthipphavong, V., Ali, F., Lozano, M., & Kwan, A. C. (2016). Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior. *Nature Neuroscience*, *19*(9), 1234–1242. https://doi.org/10.1038/nn.4342

Siniscalchi, M. J., Wang, H., & Kwan, A. C. (2019). Enhanced Population Coding for Rewarded Choices in the Medial Frontal Cortex of the Mouse. *Cerebral Cortex*, *29*(10), 4090–4106. https://doi.org/10.1093/cercor/bhy292

Sjoerds, Z., Brink, W. van den, Beekman, A. T. F., Penninx, B. W. J. H., & Veltman, D. J. (2014). Response inhibition in alcohol-dependent patients and patients with depression/anxiety: A functional magnetic resonance imaging study. *Psychological Medicine*, *44*(8), 1713–1725. https://doi.org/10.1017/S0033291713002274

Sjoerds, Z., de Wit, S., van den Brink, W., Robbins, T. W., Beekman, A. T. F., Penninx, B. W. J. H., & Veltman, D. J. (2013). Behavioral and neuroimaging evidence for overreliance on habit learning in alcohol-dependent patients. *Translational Psychiatry*, *3*, e337. https://doi.org/10.1038/tp.2013.107

Tecuapetla, F., Jin, X., Lima, S. Q., & Costa, R. M. (2016). Complementary Contributions of Striatal Projection Pathways to Action Initiation and Execution. *Cell*, *166*(3), 703–715. https://doi.org/10.1016/j.cell.2016.06.032

Yin, H. H. (2009). The role of the murine motor cortex in action duration and order. *Frontiers in Integrative Neuroscience*, *3*, 23. https://doi.org/10.3389/neuro.07.023.2009

CONCLUSION

Three sets of studies examined how individual subjective experience affected behavior and the associated neural mechanisms. In Chapter 1, the amount of experience individual mice had with a rule predicted subsequent exploitation of that rule, with OFC-M2 projection neurons involved in this experience-based exploitation. In Chapter 2, results suggested that a diverse array of experiential information was represented in and used by M2 and its DMS projections to guide behavior. In Chapter 3, prior chronic alcohol impaired behavioral flexibility specifically through its induction of hyperactive M2-DMS. Collectively, these studies make the larger point that subjective experience is not an annoyance to be factored out. Rather, it makes large contributions both to behavior and its neural control, and may help to explain and understand deficits in decision-making associated with psychiatric disease.

What do we miss by neglecting subjective experience?

Traditional approaches in neurobiological research that ignore subjective experience and average across mice and across decisions may leave us with a blinkered view of what is controlling behavior. Even if one is primarily interested in understanding, say, perception, it may still be difficult to divorce this from subjective experience given the existence of stimuli-based history-dependencies (e.g., Adam &

Serences, 2021). Indeed, different environmental interactions may also drive differences in brain structure and function in genetically identical animals (Freund et al., 2013), and in human twins (Gao et al., 2014), suggesting that averaging discrete variables across subjects may obscure differences in neural circuit recruitment.

While experimenters may have some idea of what subjects “should” do, use, or pay attention to their subjects of course do not. This presupposition of what subjects ought to do is revealed in the language used to describe history dependencies: “biases”, “lapses”, or the use of “irrelevant” information (e.g., Busse et al., 2011). Rather than asserting that subjects are making suboptimal decisions (optimality is truly in the eye of the beholder), careful behavioral analyses can reveal experience-based determinants of decision-making (e.g., Lak et al., 2020; Pisupati et al., 2021). By incorporating subjective experience in Chapters 1-3, I show that neither every mouse, nor every decision made by an individual mouse, is the same. If I had taken only traditional approaches of data analysis, averaging across mice and computing between group differences of measures such as accuracy, I would have been left with largely null results. In Chapter 1, I would have seen that temporal uncertainty during learning had no effect on the degree of explore/exploit, and missed that individual experience predicted subsequent exploitation. This would have also made it difficult to

make sense of reduced exploitation following OFC-M2 inhibition. In Chapter 2, I would have been largely at a loss to explain *how* mice performed the task, and ignorant of the many different aspects of experience used to guide performance. Furthermore, I would have been unable to see that prior reward did not dramatically affect subsequent strategy (i.e., no win-stay/lose-shift). This result challenges the prevailing dogma arising from traditional binary choice tasks, and highlights the utility of broadening our investigative techniques. In Chapters 2-3, given the lack of coarse behavioral differences, I would have concluded that M2, M2-DMS, and prior chronic alcohol all had little to no role to play in the behavior. Instead, by designing tasks that allowed for the modeling of how subjective experience contributed, I was able to show data to suggest that rodent premotor circuits integrated diverse sources of experiential information to guide flexible behavior, and that this was specifically disrupted by chronic alcohol.

From Pictures to Movies

Though they certainly provide value, 2-dimensional snapshots of the brain-behavior relationship may miss a crucial dimension; how has the relationship evolved across time and been influenced by subjective experience? Although there is a rich field of study targeting history-dependencies (e.g., Busse et al., 2011; Hattori et al.,

2019; Lak et al., 2020; Pinto et al., 2018; Pisupati et al., 2021; Siniscalchi et al., 2016, 2019), most studies still use quite constrained tasks. For instance, many involve perceptual decision-making directed towards discrete, binary options (go/no-go, respond left/right). These may not be the ideal approaches given that, as mentioned, use of experience is not *required* for performance. Indeed, that it oftentimes takes rodents and primates weeks or months of training to achieve a modicum of performance on perceptual decision-making tasks suggests that this is not a task they are well-adapted to solve, perhaps in part because they need to learn to ignore their experience in favor of attending only to the current stimuli. Such tasks may also introduce conflict between prior and new learning through the need for involved shaping procedures, and prior experience can affect subsequent learning/performance and the involved neural circuitry (Gire et al., 2016; Jacob et al., 2021; Sharpe et al., 2021). In contrast, many animals reached maximal performance in the lever hold down task after only a few days of training in Chapters 2-3. This suggests that researchers, in addition to examining subjective experience in existing tasks, ought to design and tweak tasks with the consideration that animals might be strongly adapted to use this sort of information. This is analogous to the notion that researchers should consider the “umwelt” or the unique perceptual world of an organism (Von Uexküll, 1934) when investigating animal perception (Caves et al., 2019).

The natural world is not chopped up into discrete trials or bins; one moment flows irrevocably into the next. Thus, though useful for analyses, it may be problematic to truncate tasks into trials and bins and assume that no relevant computations occur during intertrial periods. Consistent, static cues delineating discrete choices to be made or withheld may also be relatively rare in the non-stationary natural world, and there is a risk that the use of such cues could lead to elicited behavior. This is particularly noteworthy since voluntary and elicited behaviors can recruit distinct neural circuits, including primate Pre-SMA/SMA which is recruited in voluntary, but not cued movement (Okano & Tanji, 1987; Thaler et al., 1995). The all-or-nothing nature of discrete variables may also make it difficult to understand how experience alters behavior in a more continuous fashion (e.g., alterations in action execution or timing). Thus, while we may wish to study how some specific process or computation is implemented in the brain, this implementation may be context and experience-dependent (Bouton & Balleine, 2019; Jacob et al., 2021; Sharpe et al., 2021).

A series of recent studies have reported remarkably widespread encoding of sensation, movement, and decision-variables in the brain (Allen et al., 2017, 2019; Musall et al., 2019; Peters et al., 2021; Pinto et al., 2019; Steinmetz et al., 2019). This has sparked a number of interpretations (e.g., distributed motor commands, re-

afferents, predictive processing (Kaplan & Zimmer, 2020)). However, by averaging across discrete decisions and animals without regard to subjective experience, these approaches may essentially provide us with a 2-dimensional snapshot of the brain-behavior relationship. Based purely on such studies, it may be difficult to say if/how any neural circuit uniquely contributes to neural computation (including M2, which is one of the many regions identified in the above studies). Given that the perceptual decision-making tasks used in these studies show history dependencies (Busse et al., 2011; Lak et al., 2020; Pinto et al., 2018), this ought to be reflected in the brain. In fact, widefield calcium recordings have revealed that many dorsal cortical regions appear to be modulated by history (Hattori et al., 2019). If we ignore the different behavioral determinants of individual decisions in individual animals, we essentially neglect an entire dimension of data; a dimension likely to have extreme relevance for adaptive behavior and human disease.

Subjective experience in psychiatric disease

Psychiatric disease research could benefit from a renewed focus on subjective experience. Human disease exists in a fully open, continuously evolving world, where experiential effects are likely to play a role. Indeed, the pattern of drug consumption can prove pivotal in the etiology of a substance use disorder (Allain et al., 2015). Thus,

modeling the behavioral controllers of decisions concerning “when, how, and how much” may be informative. Additionally, there has been a drive to categorize disrupted decision-making in psychiatric disorders, including the concepts of behavioral inflexibility, impulsivity, compulsivity, and habits (Everitt & Robbins, 2005, 2016; Gillan et al., 2011; Graybiel & Rauch, 2000; Winstanley et al., 2010). While these are useful concepts (especially in diagnosis), each of them is supported by a plethora of different computations. Investigation of how specific aspects of subjective experience are affected in psychiatric disease may be essential for mechanistic insight to inform novel treatment development, especially since these different computations are supported by different neural mechanisms (Everitt & Robbins, 2005, 2016; Lerner, 2020; Yin & Knowlton, 2006). Chapter 3 shows the possible utility of this approach, as results suggest that prior chronic alcohol specifically reduces sensitivity to recent experience, and does so via long-lasting neuroadaptations that cause hyperactivity in M2-DMS projection neurons. This may provide a mechanistic explanation for previously observed deficits in behavioral flexibility associated with AUD (Claus et al., 2011; Duka et al., 2011; Scaife & Duka, 2009; Shnitko et al., 2020; Sjoerds et al., 2014), and suggests human premotor regions as a novel target for AUD treatment.

How does M2 contribute to experience-based decision-making?

Chapters 1-3 show that, at least in the case of M2, subjective experience is a powerful driver of neural circuit activity and recruitment. While M2 has been proposed to be involved in sensory-based decision-making (Barthas & Kwan, 2017), the present, as well as prior works (Hattori et al., 2019; Murakami et al., 2017; Siniscalchi et al., 2016, 2019; Sul et al., 2011), suggest a broader role for M2 in using subjective experience - effectively integrating perceptual, experiential, and internal information - to guide decision-making and develop experience-based motor plans.

Given this fairly broad hypothesis, it is important to note what M2 is *not* needed for. As we saw in Chapters 1-3, M2 lesion or inhibition generally had no obvious effect on the acquisition of coarse behavioral measures, consistent with prior literature (Gremel & Costa, 2013). While M2 seems to be needed for the integration of subjective experience to modify behavior, not all behavior necessarily needs this integration and/or fine-grained modification. Pavlovian or habitual learning are sufficient to acquire many behaviors (Yin and Knowlton, 2006). This highlights the problem with treating decisions as interchangeable, since careful investigation is required to understand the behavioral controllers of decision-making (Balleine, 2019). While prior works suggest that habitual strategies may control behavior with M2 offline (Gremel & Costa, 2013; Sul et al., 2011), Chapter 2 reveals a specific deficit that may

be responsible for habitual control; an inability to integrate different aspects of experience to appropriately adapt behavior.

How does M2 integrate subjective experience? Separate populations within M2 itself have the capability to provide both the input and output of an integrator (Murakami et al., 2014). On the other hand, M2 likely also receives important sensory, associative, and historical information from other neural circuits (Reep et al., 1987, 1990; Zingg et al., 2014). Chapter 1 suggests OFC-M2 projections are one such circuit, serving as a crucial mediator of rule learning to control the relative balance of exploit (i.e., experience-based) versus explore decisions. Although I have argued that M2 is involved in using subjective experience, given the behavioral relevance of subjective experience, it is unlikely to be the *only* such neural circuit. For instance, retrosplenial cortex (RSC) also represents historical information, and in fact, relative to M2 its activity can more accurately decode historical information (Hattori et al., 2019). M2 and RSC cortex are reciprocally connected (Yamawaki et al., 2016), though it is unknown what role (if any) these projections play. It may be that RSC-M2 projections convey information about the history of rewarded actions, and this is integrated with other information sources (e.g., posterior parietal cortex (Hwang et al., 2021)) to plan, select, or modify actions based on experience.

After integrating experiential information, M2 may convey these experience-based motor plans to downstream structures such as dorsal striatum. In Chapters 2-3, M2-DMS projections were specifically needed to implement a recent experience-based strategy. This also helps to explain the role of M2 (Yin, 2009) and M2-dorsal striatal projections in sequence learning (Rothwell et al., 2015), as proper sequence execution requires that animals respond appropriately based on recent experience (e.g., after action A make action B). Of particular note, in Chapters 2-3 pre-training lesions, chemogenetic inhibition, and optogenetic inhibition all induced similar effects, suggesting that M2-DMS plays an instructive role, i.e., it provides information/computations not otherwise available (Otchy et al., 2015). Additionally, this instructive role seems to be restricted to *during* performance itself (Chapter 2), rather than before or after. This is somewhat suggestive of a role for M2-DMS as a comparator – comparing current actions with prior ones – as has been hypothesized to occur in premotor regions in avian vocal learning (Mooney, 2009). This also sheds some light on the fact that, while M2 is not needed to acquire simple lever press behaviors, it *is* required for learning some complex motor skills (Cao et al., 2015; Kawai et al., 2015; Makino et al., 2017), where such a comparator process is likely to be essential. If M2-DMS is functioning as a comparator, any disruption to the fidelity of the comparison should affect performance. This is supported by one of the more

intriguing results of Chapters 2-3; increases (prior alcohol), decreases (chemogenetic or optogenetic inhibition), and disruptions (lesion) in M2-DMS activity all led to similar behavioral deficits – namely, a reduced sensitivity to recent experience. This suggests that there is an optimal level or pattern of M2-DMS activity such that increases or decreases (or lesion) impairs the faithfulness of M2's representation/integration of prior experience, the comparison function itself, or communication to downstream regions. However, at the moment such a role remains speculative. Future work – particularly at the level of single cells – would be invaluable to determine if M2/M2-DMS does function as a comparator. Additionally, though some work has been done (Emmons et al., 2017; Rothwell et al., 2015; Vargo and Marshall, 1995), examination of the downstream consequences of changes in M2 activity would also be of interest, particularly in disease states like AUD (Chapter 3), OCD (Corbit et al., 2019), or Parkinson's Disease (Magno et al., 2019).

Conclusion

Though there is a clear trade-off between relatively unconstrained/naturalistic versus controlled laboratory approaches (Juavinett et al., 2018), I believe we have drifted somewhat too far in the direction of control, and join others in calling for a renewed focus on behavior (Gomez-Marin et al., 2014; Gomez-Marin & Ghazanfar,

2019; Krakauer et al., 2017; Schreiner et al., 2021; Yoo et al., 2021). We need not shift too greatly; in this dissertation, just by using relatively unconstrained tasks without trials, cues, restraint, or discrete choices, I found data to suggest that mice can use a wealth of individually experienced information, including internal state, time, and checking behavior to guide their decision-making. Subjective experience powerfully affected the activity in, and was controlled by, premotor corticostriatal circuits. These premotor circuits integrated diverse experiential sources to bias strategy-level decisions, and this computation was specifically impacted by prior chronic alcohol to impair behavioral flexibility.

Greater investigation of subjective experience may prove to be a generally useful approach. Although most psychological and neurobiological investigations do not seek to model how these aspects of subjective experience are affecting their data, they likely *are*. Ironically, by constraining and limiting our investigations to increase control, specificity, and replicability, we may have occluded our ability to reveal fundamental and generalizable mechanisms.

Heinrich Schliemann was a 19th century archaeologist famous for his excavation of the ancient city of Troy. Although Schliemann claimed to have found the Troy of Homer's *The Iliad*, as well as "King Priam's Treasure", subsequent analysis

revealed that Schliemann's Troy and "Priam's Treasure" were much too old (Easton, 1998). Unfortunately, Schliemann's hasty digging and single-minded pursuit of Homer's Troy not only neglected important contextual details, it also significantly damaged the more superficial layers, which were *actually* the ruins of Troy. Looking at the current neurobiological focus on constraining behavior to identify specific task-related variables, I am left wondering; what treasures are *we* misidentifying – or missing entirely?

References

Adam, K. C. S., & Serences, J. T. (2021). History modulates early sensory processing of salient distractors. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3099-20.2021

Allain, F., Minogianis, E.-A., Roberts, D. C. S., & Samaha, A.-N. (2015). How fast and how often: The pharmacokinetics of drug use are decisive in addiction. *Neuroscience & Biobehavioral Reviews*, *56*, 166–179. https://doi.org/10.1016/j.neubiorev.2015.06.012

Allen, W. E., Chen, M. Z., Pichamoorthy, N., Tien, R. H., Pachitariu, M., Luo, L., & Deisseroth, K. (2019). Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science*, eaav3932. https://doi.org/10.1126/science.aav3932

Allen, W. E., Kauvar, I. V., Chen, M. Z., Richman, E. B., Yang, S. J., Chan, K., Gradinaru, V., Deverman, B. E., Luo, L., & Deisseroth, K. (2017). Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron*, *94*(4), 891-907.e6. https://doi.org/10.1016/j.neuron.2017.04.017

Balleine, B. W. (2019). The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. *Neuron*, *104*(1), 47–62. https://doi.org/10.1016/j.neuron.2019.09.024

Barthas, F., & Kwan, A. C. (2017). Secondary Motor Cortex: Where 'Sensory' Meets 'Motor' in the Rodent Frontal Cortex. *Trends in Neurosciences*, *40*(3), 181–193. https://doi.org/10.1016/j.tins.2016.11.006

Bouton, M. E., & Balleine, B. W. (2019). Prediction and control of operant behavior: What you see is not all there is. *Behavior Analysis: Research and Practice*, *19*(2), 202–212. https://doi.org/10.1037/bar0000108

Busse, L., Ayaz, A., Dhruv, N. T., Katzner, S., Saleem, A. B., Schölvinck, M. L., Zaharia, A. D., & Carandini, M. (2011). The Detection of Visual Contrast in the Behaving Mouse. *The Journal of Neuroscience*, *31*(31), 11351–11361. https://doi.org/10.1523/JNEUROSCI.6689-10.2011

Cao, V. Y., Ye, Y., Mastwal, S., Ren, M., Coon, M., Liu, Q., Costa, R. M., & Wang, K. H. (2015). Motor Learning Consolidates Arc-Expressing Neuronal Ensembles in Secondary Motor Cortex. *Neuron*, *86*(6), 1385–1392. https://doi.org/10.1016/j.neuron.2015.05.022

Caves, E. M., Nowicki, S., & Johnsen, S. (2019). Von Uexküll Revisited: Addressing Human Biases in the Study of Animal Perception. *Integrative and Comparative Biology*, *59*(6), 1451–1462. https://doi.org/10.1093/icb/icz073

Claus, E. D., Kiehl, K. A., & Hutchison, K. E. (2011). Neural and Behavioral Mechanisms of Impulsive Choice in Alcohol Use Disorder. *Alcoholism, Clinical and Experimental Research*, *35*(7), 1209–1219. https://doi.org/10.1111/j.1530-0277.2011.01455.x

Corbit, V. L., Manning, E. E., Gittis, A. H., & Ahmari, S. E. (2019). Strengthened Inputs from Secondary Motor Cortex to Striatum in a Mouse Model of Compulsive Behavior. *Journal of Neuroscience*, *39*(15), 2965–2975. https://doi.org/10.1523/JNEUROSCI.1728-18.2018

Duka, T., Trick, L., Nikolaou, K., Gray, M. A., Kempton, M. J., Williams, H., Williams, S. C. R., Critchley, H. D., & Stephens, D. N. (2011). Unique Brain Areas Associated with Abstinence Control Are Damaged in Multiply Detoxified Alcoholics. *Biological Psychiatry*, *70*(6), 545–552. https://doi.org/10.1016/j.biopsych.2011.04.006

Easton, D. F. (1998). Heinrich Schliemann: Hero or Fraud? *The Classical World*, *91*(5), 335–343. https://doi.org/10.2307/4352102

Emmons, E. B., Corte, B. J. D., Kim, Y., Parker, K. L., Matell, M. S., & Narayanan, N. S. (2017). Rodent Medial Frontal Control of Temporal Processing in the Dorsomedial Striatum. *Journal of Neuroscience*, *37*(36), 8718–8733. https://doi.org/10.1523/JNEUROSCI.1376-17.2017

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience*, *8*(11), 1481–1489. https://doi.org/10.1038/nn1579

Everitt, B. J., & Robbins, T. W. (2016). Drug Addiction: Updating Actions to Habits to Compulsions Ten Years On. *Annual Review of Psychology*, *67*(1), 23–50. https://doi.org/10.1146/annurev-psych-122414-033457

Freund, J., Brandmaier, A. M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., Sachser, N., Lindenberger, U., & Kempermann, G. (2013). Emergence of Individuality in Genetically Identical Mice. *Science*, *340*(6133), 756–759. https://doi.org/10.1126/science.1235294

Gao, W., Elton, A., Zhu, H., Alcauter, S., Smith, J. K., Gilmore, J. H., & Lin, W. (2014). Intersubject Variability of and Genetic Effects on the Brain's Functional Connectivity during Infancy. *Journal of Neuroscience*, *34*(34), 11288–11296. https://doi.org/10.1523/JNEUROSCI.5072-13.2014

Gillan, C. M., Papmeyer, M., Morein-Zamir, S., Sahakian, B. J., Fineberg, N. A., Robbins, T. W., & de Wit, S. (2011). Disruption in the Balance Between Goal-Directed Behavior and Habit Learning in Obsessive-Compulsive Disorder. *The American Journal of Psychiatry*, *168*(7), 718–726. https://doi.org/10.1176/appi.ajp.2011.10071062

Gire, D. H., Kapoor, V., Arrighi-Allisan, A., Seminara, A., & Murthy, V. N. (2016). Mice Develop Efficient Strategies for Foraging and Navigation Using Complex Natural Stimuli. *Current Biology*, *26*(10), 1261–1273. https://doi.org/10.1016/j.cub.2016.03.040

Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M., & Mainen, Z. F. (2014). Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, *17*(11), 1455–1462. https://doi.org/10.1038/nn.3812

Gomez-Marin, A., & Ghazanfar, A. A. (2019). The Life of Behavior. *Neuron*, *104*(1), 25-36. https://doi.org/10.1016/j.neuron.2019.09.017

Graybiel, A. M., & Rauch, S. L. (2000). Toward a Neurobiology of Obsessive-Compulsive Disorder. *Neuron*, *28*(2), 343–347. https://doi.org/10.1016/S0896-6273(00)00113-6

Gremel, C. M., & Costa, R. M. (2013). Premotor cortex is critical for goal-directed actions. *Frontiers in Computational Neuroscience*, *7*. https://doi.org/10.3389/fncom.2013.00110

Hattori, R., Danskin, B., Babic, Z., Mlynaryk, N., & Komiyama, T. (2019). Area-Specificity and Plasticity of History-Dependent Value Coding During Learning. *Cell*, *177*(7), 1858-1872.e15. https://doi.org/10.1016/j.cell.2019.04.027

Hwang, E. J., Dahlen, J. E., Mukundan, M., & Komiyama, T. (2021). Disengagement of Motor Cortex during Long-Term Learning Tracks the Performance Level of Learned Movements. *Journal of Neuroscience*, *41*(33), 7029–7047. https://doi.org/10.1523/JNEUROSCI.3049-20.2021

Jacob, P. F., Vargas-Gutierrez, P., Okray, Z., Vietti-Michelina, S., Felsenberg, J., & Waddell, S. (2021). Prior experience conditionally inhibits the expression of new learning in Drosophila. *Current Biology*, *31*(16), 3490-3503.e3. https://doi.org/10.1016/j.cub.2021.05.056

Juavinett, A. L., Erlich, J. C., & Churchland, A. K. (2018). Decision-making behaviors: Weighing ethology, complexity, and sensorimotor compatibility. *Current Opinion in Neurobiology*, *49*, 42–50. https://doi.org/10.1016/j.conb.2017.11.001

Kaplan, H. S., & Zimmer, M. (2020). Brain-wide representations of ongoing behavior: A universal principle? *Current Opinion in Neurobiology*, *64*, 60–69. https://doi.org/10.1016/j.conb.2020.02.008

Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A. L., Dhawale, A. K., Kampff, A. R., & Ölveczky, B. P. (2015). Motor Cortex Is Required for Learning but Not for Executing a Motor Skill. *Neuron*, *86*(3), 800–812. https://doi.org/10.1016/j.neuron.2015.03.024

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Lak, A., Hueske, E., Hirokawa, J., Masset, P., Ott, T., Urai, A. E., Donner, T. H., Carandini, M., Tonegawa, S., Uchida, N., & Kepecs, A. (2020). Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. *ELife*, *9*, e49834. https://doi.org/10.7554/eLife.49834

Lerner, T. N. (2020). Interfacing behavioral and neural circuit models for habit formation. *Journal of Neuroscience Research*, *98*(6), 1031–1045. https://doi.org/10.1002/jnr.24581

Magno, L. A. V., Tenza-Ferrer, H., Collodetti, M., Aguiar, M. F. G., Rodrigues, A. P. C., Silva, R. S. da, Silva, J. do P., Nicolau, N. F., Rosa, D. V. F., Birbrair, A., Miranda, D. M., & Romano-Silva, M. A. (2019). Optogenetic stimulation of the M2 cortex reverts motor dysfunction in a mouse model of Parkinson's Disease. *Journal of Neuroscience*, *39*(17), 3234–3248. https://doi.org/10.1523/JNEUROSCI.2277-18.2019

Makino, H., Ren, C., Liu, H., Kim, A. N., Kondapaneni, N., Liu, X., Kuzum, D., & Komiyama, T. (2017). Transformation of Cortex-wide Emergent Properties during Motor Learning. *Neuron*, *94*(4), 880-890.e8. https://doi.org/10.1016/j.neuron.2017.04.015

Mooney, R. (2009). Neural mechanisms for learned birdsong. *Learning & Memory*, *16*(11), 655–669. https://doi.org/10.1101/lm.1065209

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

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

Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., & Churchland, A. K. (2019). Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, *22*(10), 1677–1686. https://doi.org/10.1038/s41593-019-0502-4

Okano, K., & Tanji, J. (1987). Neuronal activities in the primate motor fields of the agranular frontal cortex preceding visually triggered and self-paced movement. *Experimental Brain Research*, *66*(1), 155–166. https://doi.org/10.1007/BF00236211

Otchy, T. M., Wolff, S. B. E., Rhee, J. Y., Pehlevan, C., Kawai, R., Kempf, A., Gobes, S. M. H., & Ölveczky, B. P. (2015). Acute off-target effects of neural circuit manipulations. *Nature*, *528*(7582), 358–363. https://doi.org/10.1038/nature16442

Peters, A. J., Fabre, J. M. J., Steinmetz, N. A., Harris, K. D., & Carandini, M. (2021). Striatal activity topographically reflects cortical activity. *Nature*, *591*(7850), 420–425. https://doi.org/10.1038/s41586-020-03166-8

Pinto, L., Koay, S. A., Engelhard, B., Yoon, A. M., Deverett, B., Thiberge, S. Y., Witten, I. B., Tank, D. W., & Brody, C. D. (2018). An Accumulation-of-Evidence Task Using Visual Pulses for Mice Navigating in Virtual Reality. *Frontiers in Behavioral Neuroscience*, *12*, 36. https://doi.org/10.3389/fnbeh.2018.00036

Pinto, L., Rajan, K., DePasquale, B., Thiberge, S. Y., Tank, D. W., & Brody, C. D. (2019). Task-Dependent Changes in the Large-Scale Dynamics and Necessity of Cortical Regions. *Neuron*, *104*(4), 810-824.e9. https://doi.org/10.1016/j.neuron.2019.08.025

Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., & Churchland, A. K. (2021). Lapses in perceptual decisions reflect exploration. *ELife*, *10*, e55490. https://doi.org/10.7554/eLife.55490

Reep, R. L., Corwin, J. V., Hashimoto, A., & Watson, R. T. (1987). Efferent Connections of the Rostral Portion of Medial Agranular Cortex in Rats. *Brain Research Bulletin*, *19*(2), 203–221. https://doi.org/10.1016/0361-9230(87)90086-4

Reep, R. L., Goodwin, G. S., & Corwin, J. V. (1990). Topographic organization in the corticocortical connections of medial agranular cortex in rats. *The Journal of Comparative Neurology*, *294*(2), 262–280. https://doi.org/10.1002/cne.902940210

Rothwell, P. E., Hayton, S. J., Sun, G. L., Fuccillo, M. V., Lim, B. K., & Malenka, R. C. (2015). Input- and Output-Specific Regulation of Serial Order Performance by Corticostriatal Circuits. *Neuron*, *88*(2), 345–356. https://doi.org/10.1016/j.neuron.2015.09.035

Scaife, J. C., & Duka, T. (2009). Behavioural measures of frontal lobe function in a population of young social drinkers with binge drinking pattern. *Pharmacology, Biochemistry, and Behavior*, *93*(3), 354–362. https://doi.org/10.1016/j.pbb.2009.05.015

Sharpe, M. J., Batchelor, H. M., Mueller, L. E., Gardner, M. P. H., & Schoenbaum, G. (2021). Past experience shapes the neural circuits recruited for future learning. *Nature Neuroscience*, *24*(3), 391–400. https://doi.org/10.1038/s41593-020-00791-4

Shnitko, T. A., Gonzales, S. W., Newman, N., & Grant, K. A. (2020). Behavioral Flexibility in alcohol drinking monkeys: The morning after. *Alcoholism, Clinical and Experimental Research*, *44*(3), 729–737. https://doi.org/10.1111/acer.14289

Schreiner, D. C., Yalcinbas, E. A., & Gremel, C. M. (2021). A push for examining subjective experience in value-based decision-making. *Current Opinion in Behavioral Sciences*, *41*, 45–49. https://doi.org/10.1016/j.cobeha.2021.03.020

Siniscalchi, M. J., Phoumthipphavong, V., Ali, F., Lozano, M., & Kwan, A. C. (2016). Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior. *Nature Neuroscience*, *19*(9), 1234–1242. https://doi.org/10.1038/nn.4342

Siniscalchi, M. J., Wang, H., & Kwan, A. C. (2019). Enhanced Population Coding for Rewarded Choices in the Medial Frontal Cortex of the Mouse. *Cerebral Cortex*, *29*(10), 4090–4106. https://doi.org/10.1093/cercor/bhy292

Sjoerds, Z., Brink, W. van den, Beekman, A. T. F., Penninx, B. W. J. H., & Veltman, D. J. (2014). Response inhibition in alcohol-dependent patients and patients with depression/anxiety: A functional magnetic resonance imaging study. *Psychological Medicine*, *44*(8), 1713–1725. https://doi.org/10.1017/S0033291713002274

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 1–8. https://doi.org/10.1038/s41586-019-1787-x

Sul, J. H., Jo, S., Lee, D., & Jung, M. W. (2011). Role of rodent secondary motor cortex in value-based action selection. *Nature Neuroscience*, *14*(9), 1202–1208. https://doi.org/10.1038/nn.2881

Thaler, D., Chen, Y. C., Nixon, P. D., Stern, C. E., & Passingham, R. E. (1995). The functions of the medial premotor cortex. I. Simple learned movements. *Experimental Brain Research*, *102*(3), 445–460.

Vargo, J. M., & Marshall, J. F. (1995). Time-dependent changes in dopamine agonist-induced striatal fos immunoreactivity are related to sensory neglect and its recovery after unilateral prefkontal cortex injury. *Synapse*, *20*(4), 305–315. https://doi.org/10.1002/syn.890200404

Von Uexküll, J. (1934). A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*, *89*(4). https://doi.org/10.1515/semi.1992.89.4.319

Winstanley, C. A., Olausson, P., Taylor, J. R., & Jentsch, J. D. (2010). Insight Into the Relationship Between Impulsivity and Substance Abuse From Studies Using

Animal Models. *Alcoholism: Clinical and Experimental Research*, *34*(8), 1306–1318. https://doi.org/10.1111/j.1530-0277.2010.01215.x

Yamawaki, N., Radulovic, J., & Shepherd, G. M. G. (2016). A Corticocortical Circuit Directly Links Retrosplenial Cortex to M2 in the Mouse. *The Journal of Neuroscience*, *36*(36), 9365–9374. https://doi.org/10.1523/JNEUROSCI.1099-16.2016

Yin, H. H. (2009). The role of the murine motor cortex in action duration and order. *Frontiers in Integrative Neuroscience*, *3*, 23. https://doi.org/10.3389/neuro.07.023.2009

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476. https://doi.org/10.1038/nrn1919

Yoo, S. B. M., Hayden, B. Y., & Pearson, J. M. (2021). Continuous decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1819), 20190664. https://doi.org/10.1098/rstb.2019.0664

Zingg, B., Hintiryan, H., Gou, L., Song, M. Y., Bay, M., Bienkowski, M. S., Foster, N. N., Yamashita, S., Bowman, I., Toga, A. W., & Dong, H.-W. (2014). Neural Networks of the Mouse Neocortex. *Cell*, *156*(5), 1096–1111. https://doi.org/10.1016/j.cell.2014.02.023