Beyond Simple Tests of Value:

A neuroeconomic, translational, disease-relevant, and circuit-based approach to resolve the computational complexity of decision making

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Brian Musa Sweis
M.D./Ph.D. Candidate

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Doctor of Philosophy

ADVISED BY

Mark J. Thomas, Ph.D. & A. David Redish, Ph.D.

July 2018

Acknowledgements & Dedication

I would like to thank a number of individuals; for, without their invaluable support, none of this – either the body of work contained in this thesis or my professional and personal development – would be possible.

I am indebted to the past and present administrative directors and supporting staff of the University of Minnesota's Medical Scientist (MD/PhD) Training Program and Graduate Program in Neuroscience, as well as several departments and centers including the Neuroscience Department, Psychology Department, Department of Integrative Biology and Physiology, and Center for Cognitive Sciences. Included in these teams, I would like to individually thank Dr. Yoji Shimizu, Susan Shurson, Nicholas Berg, Dr. Marshall Hertz, Dr. Lisa Schimmenti, Dr. Bryce Binstadt, Dr. Linda McLoon, John Paton, Elaine McCauley, Lateeph Onikoro, Dr. Dori Henderson, Kirsti Hendricksen, and Cindy Marceau.

I would also like to acknowledge my early mentors in science at Loyola University Chicago, including Dr. Louis Lucas, Dr. Robert Morrison, and Dr. Rebecca Silton. Thank you for opening my eyes to science, grounding me, cautioning me on all that might lie on the path ahead (I get it now), and encouraging me to never settle.

I thank my mentors during my early stages at the University of Minnesota, including Dr. William Engeland, Dr. Alessandro Bartolomucci, Dr. Michael Benneyworth, Dr. Afshin Divani, and others whose I felt nearly took me in as an extension of their own labs (even though I technically wasn't, Dr. Jonathan Gewirtz and Dr. John Osborn).

Individuals in the clinical realms of my education, I also thank you. Dr. Jerrold Vitek, Dr. Sophia Vinogradov, Dr. Miguel Fiol, Dr. William Orr, Dr. Jose Pardo, Dr. Mateo Calderon-Arnulphi, and especially Dr. Ezgi Tiryaki, who taught me that compassion as a healthcare provider is as every bit as if not more important than scientific inquiry.

I want to thank Dr. Michael Georgieff for his career advice, even as early as my first interview with him at the University of Minnesota, welcoming and resonating with my academic interests and pointing me in the direction of Dr. Mark Thomas and Dr. David Redish.

Of greatest importance, I cannot begin to express my gratitude for having worked with, who I think, are two of the University of Minnesota's rock stars – my thesis advisers and friends, Dr. Mark Thomas and Dr. David Redish. You two have made my experience in science incredibly exciting, incredibly valuable, and incredibly tough. Thank you for providing me with the right level of guidance and autonomy, for pushing me and putting up with me, and letting me do the same to you / with you. Your mentorship has equipped me with a sharpness that will be an asset for the rest of my life. Thank you for keeping me on my toes.

I want to thank all members of the Thomas and Redish labs, past, present, and future, including members of collaborating labs I have had the pleasure to work with, including Dr. Angus MacDonald.

I want to acknowledge the other members of my thesis committee: Dr. Matthew Chafee (chair), Dr. Patrick Rothwell (member), and Dr. Kelvin Lim (outside, physician-scientist).

I want to acknowledge Dr. Geoff Ghose for his mentorship in teaching over the years. I want to thank Jim Beattie, who has also instructed me in the art of teaching, and whose friendship and lessons I cherish.

Lastly, I want to dedicate my thesis to my family: Musa, Majedah, Marvet (+Demijan), Sam (+Katie), Auddie (+Tina), Jamie, 'nistian, Lainey, and Obi-Wan Ben Kenobi, as well as the extended Sweis network. Thank you for tolerating listening to my science as long as you do. I hope more than anything I make you all in Chicago proud.

My efforts and the work presented in this thesis were partially supported by the National Institutes of Health, including two program training grants through the National Institute of General Medical Sciences (NIGMS T32GM008244, NIGMS T32GM008471) and my individually funded research fellowships, including the Ruth L. Kirschstein National Research Service Award (NRSA) through the National Institute on Drug Abuse (NIDA F30 DA043326) and the MnDRIVE Neuromodulation Research Fellowships. This support is in addition to funding provided by my advisors' and collaborators' laboratories.

Abstract

How the brain processes information when making decisions depends on how that information is stored. Distinct neural circuits are capable of storing information in many different ways that are better suited for different situations. The decision-making processes that access those different bits of stored information are not singular and occupy separable neural circuits, each of which can operate in parallel with one another, and each of which can confer different information processing properties based on the neural constraints within which a given computation resides. Such is the framework of recent theories in neuroeconomics, which suggest that decisions are multi-faceted and action-selection processes can arise from fundamentally distinct circuit-specific neural computations. In this thesis, I present a body of work that takes a neuroeconomics approach through a series of experiments that reveal the complexities of multiple, parallel decision-making systems through complex behaviors by moving beyond simple tests of value. In the first half of this thesis, I demonstrate how complex behavioral computations can resolve fundamentally distinct valuation algorithms thought to reside in separable neural circuits. I then translate this approach between human and non-human rodent animal models in order to reveal how multiple, parallel decision-making systems are conserved across species over evolution. In the second half of this thesis, I demonstrate the utility of behavioral economics in disease-relevant and circuit-based studies. If multiple, parallel decision-making processes are thought to be intimately related to the heterogeneous ways in which information can be stored in separable neural circuits, I examine how addiction – a disease which is thought to be a disorder of the neurobiological mechanisms of learning and memory – might alter how stored information is processed in separable decision-making systems uniquely using a mouse model of two different forms of addiction. In doing so, I demonstrate how different forms of addiction give rise to unique, lasting vulnerabilities in fundamentally distinct decision-making computations. These discoveries can aid in resolving neuropsychiatric disease heterogeneity by moving beyond simple tests of value where complex behaviors that are measured can more accurately reflect the neurally distinct computations that underlie those behaviors. Finally, I take a neuromodulation approach and directly alter the strength of synaptic transmission in a circuit-specific manner using optogenetics in mice tested in this neuroeconomic framework. I demonstrate how plasticity alterations in projections between the infralimbic cortex and the nucleus accumbens are capable of giving rise to long-lasting disruptions of self-control related decision processes in a foraging valuation algorithm independent of and separate from a deliberative valuation algorithm measured within the same trial. Furthermore, I developed a novel plasticity measurement tool that is assayed at the neuronal population ensemble level and reveals individual differences in separable decision processes. The second half of the thesis demonstrates a potential biomarker to target as a circuit-computation-specific therapeutic intervention tailored to those types of decision-making dysfunctions. Taken together, I present a body of work in this thesis that demonstrates the utility of moving beyond simple tests of value in order to resolve the computational complexity of decision making.

Table of Contents

List of Tables

List of Figures

– Prologue –

Chapter 1

On neuroeconomics

Appreciating the computational complexity of decision making

Every decision we make, in one shape or form, is in service of our interactions with the world around us. Measuring how the brain computes possible actions to be selected is challenging, particularly because the information that guides decision-making processes can be stored so differently throughout the brain.

How information is stored in the brain can change how that information is accessed when decisions are being made (Redish 2013). The different ways in which the brain can store information is intimately linked to distinct decision-making mechanisms that access that stored information differently. Therefore, the neurobiological mechanisms of memory and the neurobiological mechanisms of decision-making information processing can be thought to exist as complex duals of each other.

Below, I will explore the growing literature that has been instrumental in characterizing the multiple memory systems as well as the multiple decision-making systems that reside in the brain. A critical link that bridges the concept of multiple memory systems to multiple decision-making systems has been the emerging field of neuroeconomics.

Neuroeconomics is an interdisciplinary field that seeks to describe the multifaceted ways in which the brain processes value, or the utility of selecting one action over another (Loewenstein et al. 2008). On the surface, seemingly identical decision outcomes observed through behavior could theoretically be driven by

Chapter reprinted with permissions from *Learning & Memory*, modified from:

Sweis BM, Thomas MJ, Redish AD. 2018a. Beyond simple tests of value: Measuring addiction as a heterogeneous disease of computation-specific valuation processes. *Learning & Memory* (in press).

fundamentally distinct and neurally separable valuation algorithms. Recent theories in neuroeconomics suggest that decisions made in different situations derive from different valuation functions residing in separable neural circuits (Redish 2013; Rangel et al. 2008; Loewenstein et al. 2008; Sanfey et al. 2006; Glimcher and Rustichini 2004). However, the experimental data probing such theories has been sparse.

It can be difficult to segregate parallel information processing algorithms using traditional behavioral paradigms that rely on simple tests of value. Behavioral consequences of distinct neural computations can offer appear grossly similar and thus remain unseparable. In this light, distinct neural dysfunctions would similarly result in unseparable behavioral consequences – a concept I discuss in this thesis in detail that has plagued progress of our understanding of neuropsychiatric mental illnesses.

In this thesis, I will also discuss how moving beyond simple tests of value toward approaches based on emerging theories in neuroeconomics may be able to give rise to significant advancements in the study of behavior in the same way that techniques used to study the brain have evolved over the last several decades, revolutionizing experimental approaches in neuroscience. Importantly, I discuss in this thesis how a careful approach in experimental neuroscience to probe the link between multiple memory systems and multiple decision-making systems matters and is often overlooked with the use of recently developed circuit dissection tools.

Understanding important differences in economic behavior can shape our understanding of the brain's mechanisms underlying action-selection and memory processes. In turn, neuroscientific discoveries can constrain and guide models of economics. Importantly, neuroeconomics can reveal novel insights into how decision processes in the brain, through all of its complexities and heterogeneity, might begin to malfunction in neuropsychiatric illnesses in distinct ways not previously appreciated.

Multiple memory systems in the brain

Every experience shapes our brain in one way or another. Experiences are capable of leaving anything from subtle to profoundly lasting changes in the structure and function of the brain (Milner et al. 1998). Moving

forward, this alters the way information is subsequently processed. Thus, it has also been argued that the only reason we learn and remember anything is to make better decisions (Redish and Mizumori 2015).

Any physical change in the brain that results from an experience can be considered to be a memory because such changes provide information about the historical past. Using this definition, several neuropsychiatric disorders that involve lasting changes in the brain have adopted a view of dysfunction in mechanisms of memory. For instance, addiction has been proposed to be a neurobiological disorder of learning and memory because drugs of abuse can leave lasting changes on the structure and function of the brain (Hyman 2005, Volkow 2012). These changes are thought to underlie why individuals with addiction struggle with making poor decisions. I will revisit in depth the implications of traditional and contemporary perspectives of memory and decision-making information processing for addiction later in this thesis.

There is an intimate link between memory and decision making. It can be argued that the only reason we learn and remember things is to make better decisions (Redish and Mizumori 2015). Information stored as memories within and between neural structures guide decision processes (Euston et al. 2012). Therefore, if addiction, for example, is considered to be a neurobiological disorder of learning and memory, it should also be considered a neurobiological disorder of decision-making information processing.

It is thought that humans have evolved in such a way that the brain is capable of storing information in multiple, separate memory systems each of which afford unique evolutionary advantages (Sherry and Schacter 1987). Theoretically, the existence of multiple memory systems can only afford evolutionary advantages when each system is specialized in such a way that the functional problems and environmental demands overcome by one system cannot be handled by the properties of another system, which could have been shaped by natural selection and adapted to serve other purposes (Rozin and Schull 1988). The definition of a memory system refers to interactions between separable mechanisms of information acquisition, retention, and retrieval that operate under certain rules, which may be fundamentally distinct from a separate

memory system (Sherry and Schacter 1987). Taken together, multiple separate mechanisms of memory acquisition, storage retention, and retrieval are thought to take place in neurally dissociable systems.

These principles of neurally distinct memory systems are not just limited to stages of memory formation (i.e., acquisition, storage, retrieval) but also extend to different types of information that can be acquired, stored, and retrieved. Multiple memory systems vary in terms of other properties, including the rate of learning or level of generalizability vs. specificity of stored information (O'Keefe and Nadel, 1978; Squire et al. 1993; Schacter and Tulving 1994). For instance, gradual, incremental learning involved in the acquisition of specific skills is thought to occur in a separate memory system distinct from rapid one-trial learning tied to relationships among specific episodes (Morris et al 1982; Yin et al. 2004; Tse et al 2007) or events with salient affective properties (Berridge and Robinson 1998; Dayan and Balleine 2002; Corbit and Balleine 2005). In the former example, practicing and updating repetitive motor programs over numerous trials are thought to depend on a form of reinforcement learning critically dependent on structures within the basal ganglia, including the caudate and putamen regions of the dorsal striatum (Packard and Knowlton 2002; Balleine et al 2007; Graybiel 2015). This memory system, typically referred to as procedural memory, is often spared in individuals with temporal lobe lesions that precipitate impairments in either episodic memories thought to be part of a distinct, hippocampal-dependent learning system (O'Keefe and Nadel 1978; Cohen and Squire 1980; Squire et al. 1993; Eichenbaum and Cohen 1994; Redish 1999, 2013) or emotional memories associated with specific stimuli thought to be part of an amygdala-dependent learning system (Ledoux 1998; Corbit and Balleine 2005; LeDoux and Daw 2018).

Double- and triple-dissociations between separable brain structures and multiple representational forms of memory have been demonstrated in rodents using cleverly designed behavioral paradigms where the rules or contingencies of the task require the use of different types of information stored in separable brain regions. For instance, by using rats trained on variants of a standard radial arm maze memory task that differed only in the contingencies required to successfully obtain rewards, brain region-specific lesions were capable of disrupting performance on select variants of the task but not others (McDonald and White 1993). Dorsal

striatum lesions produced deficits in win-stay contingencies, sparing performance on win-shift or cued contingencies, which were sensitive to hippocampus and amygdala lesions, respectively (McDonald and White 1993). Similarly, in rats trained on a standard T-maze memory task, hippocampal vs. dorsal striatum lesions differentially affect performance depending on the degree to which animals were trained. Prolonged training under regular contingencies rendered behavior no-longer sensitive to hippocampal lesions but instead sensitive to dorsal striatum lesions (Tolman 1948, Hull 1952; Packard and McGaugh 1992; Schmidt et al 2013; Gardner et al 2013). Taken together, the acquisition and expression of certain types of memories appear to take place in neurally separable learning and memory systems that differ depending on a number of properties of that stored information (O'Keefe and Nadel 1978; Squire et al. 1993; Schacter and Tulving 1994; Redish 1999, 2013).

Multiple decision-making systems in the brain

Just as separable memory systems are capable of storing different types of information, neurally dissociable decision-making systems exist to access those separate aspects of stored information. How data is stored can change how it is processed during decision-making. There is a tight relationship between the multiple representational forms that underlie memory and the multiple action-selection systems that are in play when accessing that stored information. Properties that govern differences in the cellular mechanism of storage, rate of learning acquisition, degree of information distribution across cells, and the different circuit networks within which these processes take place can confer differences in how that stored information is accessed.

Multiple decision-making systems can operate in parallel with one another and provide tradeoffs between decision properties, such as speed of processing, depth of planning, degree of flexibility, and a diversity of other factors that can influence choice (van der Meer et al. 2012). Multiple decision-making systems, which can be updated through unique forms of learning and are thought to reside in separable neural circuits, are thought to have evolved over time because each can be better suited for different situations (O'Keefe and Nadel 1978; Hikosaka et al 1999; Doya 1999; Daw et al. 2005; Rangel et al. 2008; Redish 2013).

Recent theories in neuroeconomics suggest that complex decisions are multi-faceted and reward valuations can arise from dissociable computations in distinct neural circuits (van der Meer 2012; Loewenstein et al.

2008; Rangel et al. 2008). For instance, decisions driven by emotion, decisions planned out after extended deliberation, and decisions made from practiced habit, each arise from dissociable neural processes dependent on different neural circuits. This concept is similar to the fact that multiple memory systems uniquely related to each of these three previous examples can exist in the brain, but importantly differs in that such separable decision processes can gain access to these different types of memories simultaneously and in parallel with one another during on-going behaviors. Thus, carefully-designed behavioral tasks are required to elucidate how multiple, parallel decision-making systems work together or compete with one another in order to access separable memories and drive behavior in the moment.

Pavlovian action-selection systems entail genetically hardwired motivational state-response action pairs that are capable of being associated with predictive stimuli through conditioning (Clark et al. 2012; Dayan and Berridge 2014). Physiological states are capable of driving motivation (e.g., hunger) and are linked to unconditioned responses (e.g., salivating). Importantly, these processes can be directly transferred to informative cues in the world. For example, images that are associated with a certain reward, rather than simply predicting upcoming reward availability or opportunities, can themselves adopt intrinsic value. This concept, termed incentive salience, can trigger feelings of wanting or craving in response to cue presentation and promote reward-seeking behaviors (Robinson and Berridge 1993; Bernheim and Rangel 2004; Berridge and Robinson 2016). The role of amygdala-related circuitry has been heavily implicated in these mechanisms (Clark et al. 2012; Wassum and Izquierdo 2015). Such circuits carry learned representations of sensory stimuli and integrate that information with motivational processes (LeDoux and Daw 2018). Through failure modes in these mechanisms, for example, addiction-related cues are capable of triggering decision-vulnerable states that lead to maladaptive motivated behaviors, ultimately precipitating relapse (Robinson and Flagel 2009; Walters and Redish 2018; Bernheim and Rangel 2004).

Deliberative action-selection systems entail declarative, episodic evaluation processes rooted in simulations of possible future response-outcome scenarios (Redish 2016). Deliberative valuation algorithms operate relatively slowly yet remain flexible. Hippocampus and regions of the prefrontal cortex have been heavily

implicated in these mechanisms (Johnson et al. 2007; van der Meer et al. 2012; Wang et al. 2015). Failure to engage deliberative algorithms when making decisions without planning (Everitt and Robbins 2005), a reduction in capacity to accurately simulate imaginations of possible future scenarios (Kurth-Nelson et al. 2012), or errors in the future scenarios themselves (Redish and Johnson 2007, Kurth-Nelson and Redish 2012; Goldman 1987), as well as errors in the value estimates of those future scenarios (Tiffany 2005; Naqvi and Bechara 2010) each describe fundamentally distinct and dissociable vulnerabilities in decision-making information processing within the deliberative system that may emerge in various neuropsychiatric illnesses.

Procedural action-selection systems entail well-practiced behavioral sequences that are released ballistically following the recognition of appropriate situations (Graybiel 1998, Graybiel and Grafton 2015; Redish 2013). These decision processes operate quickly yet are relatively inflexible and rely on motivational components accessed via cached value representations (Daw et al. 2005). The dorsal striatum has been heavily implicated in these mechanisms (Saint-Cyr 1999; Berke et al 2009; Gremel and Costa 2013; Smith and Graybiel 2014). Such circuits are recruited over many trials and information stored in these circuits are thought to be acquired through a form of reinforcement-like learning. Possible failure modes in the procedural system in different neuropsychiatric disorders could include increased valuation due to drug-modifications of dopaminergic signals (Redish 2004; Dezfouli et al. 2009), inabilities to extinguish perseverative motor programs (Peters et al. 2009), and strong habit-like processes that override other valuation algorithms leading to enhanced rates of information stored as procedural memories requiring less-than-usual number of training trials (Piray et al 2010).

These multiple action-selection systems, with their separate vulnerabilities, also interact, for example in the process of Pavlovian-Instrumental Transfer, in which amygdala-driven Pavlovian valuations can change the valuation stage of deliberative decisions occurring in accumbens (Talmi et al 2008; Corbitt and Balleine 2011; LeDoux and Daw 2018).

Thus, there are significant implications of the complexity of these concepts for various neuropsychiatric disorders. For instance, if addiction is to be considered a neurobiological disorder of memory, and thus decision-making, the heterogeneity with which information is both stored and processed must be taken into account. Thus, addiction-related dysfunctions in memory could be diverse and could lead to lasting heterogeneous circuit-specific changes in dissociable decision-making computational processes. These multiple vulnerabilities could generate subtly different behavioral phenotypes; however, it is also possible for distinct failure modes in separate systems to produce identical behavioral dysfunctions. Thus, seemingly identical disorders would require fundamentally distinct therapeutic interventions – a hurdle current practices in psychiatry has been struggling with overcoming.

Purpose and organization of this thesis

The purpose of this thesis is to demonstrate the utility of adopting a neuroeconomic approach to study decision making. Discoveries presented in this thesis address critical gaps in knowledge between multiple memory systems in the brain and multiple decision-making systems in the brain using a neuroeconomics approach. I present breakthroughs that illustrate how neuroeconomic approaches can be directly translated across species into human paradigms and how such translational approaches benefit from interdisciplinary collaborations across laboratories, starting with ethologically valid paradigms in non-human animals first and then translating up into human paradigms, and not the other way around, which has been commonplace in science. I present breakthroughs that reveal the utility of this approach, which can resolve neuropsychiatric disease heterogeneity using mouse models of addiction as an example case study. And I present breakthroughs that illustrate a causal relationship between mechanisms of memory and parallel mechanisms of decision-making information processing using circuit-specific neuromodulation techniques. Together, these discoveries can help the field of clinical neuropsychiatry practice begin to overcome hurdles in progress toward refining our understanding of disease etiologies and developing novel therapeutic interventions.

Measuring “value” is a daunting task and can often yield either conflicting findings depending on the task design or produce highly non-specific measurements in simple paradigms that actually reveal very little about

the neural computations that underlie those valuation. Given recent theories in neuroeconomics that suggest value is multi-faceted and can arise from multiple, parallel decision-making systems, the purpose of this thesis is to demonstrate that the multiplicity of these complex processes is conserved across species, can be carefully dissociated, even through behavior, and can reveal fundamental differences in the information processed (and valuation biases contained) in distinct decision-making systems – importantly, in a neurobiologically plausible model. In doing so, I use addiction as a case-study and demonstrate with neuroeconomics that discrete decision-making failure modes or vulnerabilities may lie in separable decision-making systems in different forms of addiction. This concept allows one to potentially resolve heterogeneity of neuropsychiatric diseases that are often masked when diagnoses or treatments are targeted to symptomology and do not address discrete etiological computational dysfunctions in the brain. I discover that such computational distinctions do indeed reside in separable neural processes that can be independently altered with circuit-specific neuromodulation.

The behavioral paradigms contained within this thesis directly build off of original work (Figure 1.1) from Adam Steiner when he was advised by David Redish (Steiner and Redish 2014) and Samantha Abram when she was co-advised by David Redish and Angus MacDonald (Abram et al. 2016). Their work has been instrumental in providing a fresh lens through which one can study behavioral neuroeconomics across species while also taking advantage of neurophysiological tools to monitor brain activity during discrete aspects of decision-making information processing.

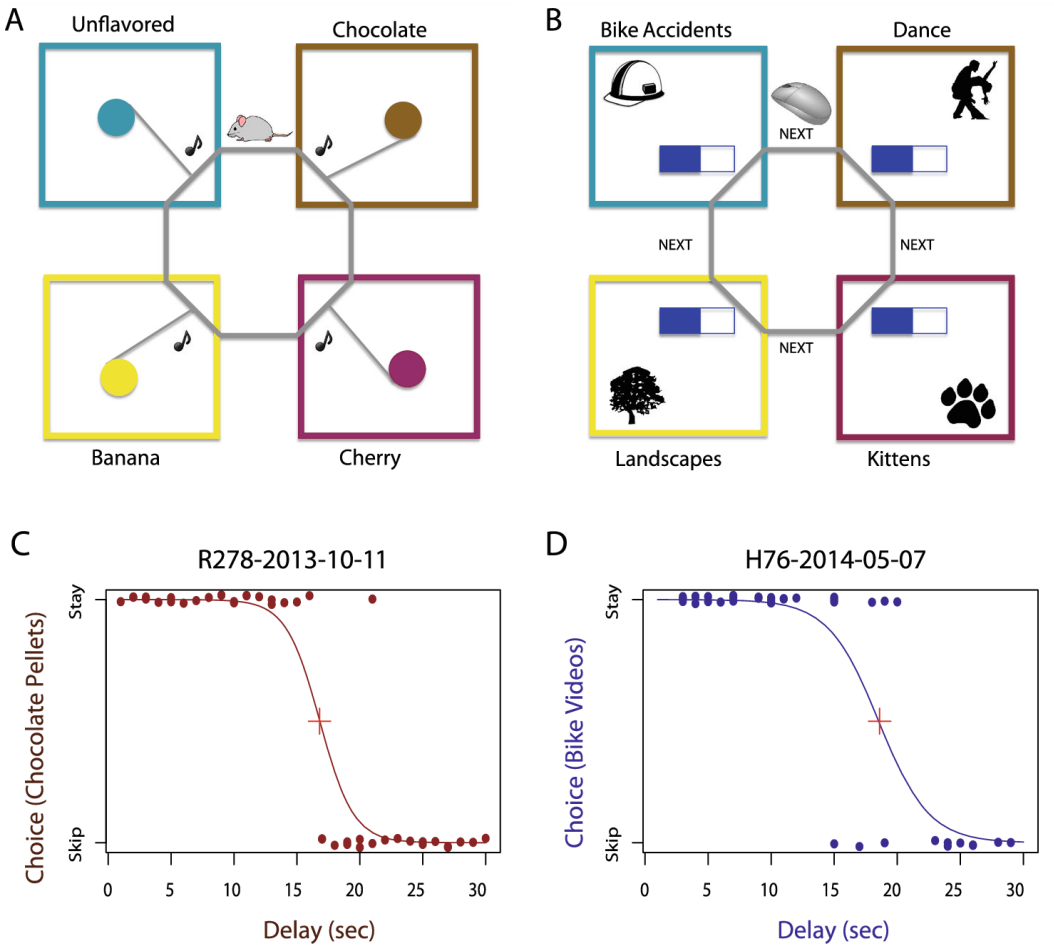
One of the hallmarks of this past research has been the construction of ultimately convincing accounts of interesting decision-making phenomena, even in non-verbal non-human animals, supported by converging behavioral and neurophysiological data. Steiner and Redish found that in special economic scenarios, when rats passed up on relatively good offers only to subsequently encounter relatively bad offers, the information gleaned at the second offer in combination with the economic violation made at the previous offer together induced regret-like behaviors in these animals (Steiner and Redish 2014). Regret was contingent on an error of one's own agency and distinct from mere disappointment. *In vivo* neural recordings in the orbitofrontal

cortex during these regret episodes revealed representations of counterfactual information processing of the missed opportunity after mistakes had been realized and that one could have made a better decision. Importantly, these regret-episodes were capable of augmenting subsequent reward valuations and could guide subsequent decisions to make up for lost efforts.

These data not only reveal the complexity of value processing in non-human animals but also provide a tractable foundation from which to interrogate a number of hypotheses regarding multiple learning and decision-making systems testable across species.

In this thesis, the remaining chapters will each tackle separate hypotheses that build off one another centered around neuroeconomic theories of multiple, parallel decision-making systems. Following each chapter that contains original research collected and analyzed during my graduate work, a short Interlogue chapter synthesizes what was learned in the broader context of neuroeconomics and multiple decision-making systems that inform and inspire the next chapter moving forward. In the middle of the thesis, based on discoveries made in earlier chapters, a new theoretical framework and perspective is discussed that delves deeper into neuropsychiatric illnesses and describes how discoveries in neuroeconomics can move computational neuropsychiatry forward, making predictions as well suggestions for the field of neuroscience as a whole, inspiring and demonstrated by the discoveries made in the final chapters of this thesis.

Figure 1.1: The original Restaurant Row (rats) and Web-Surf (humans) tasks



These figures are reproduced with permissions from Abram et al 2016 *CABN* that demonstrated the first use of the Web-Surf Task based on the Restaurant Row Task first used in Steiner and Redish in 2014 *Nature Neuroscience*. These efforts translated a rodent foraging task "back-up" into a human task, instead of the other way around. Often, analogs of human tasks that are mapped "down" onto rodent behaviors lose a number of important qualities and struggle with construct validity and face validity. (A-B) Task schematics of the two tasks. (A) In rats, a tone sounds as an animal enters a uniquely flavored "restaurant" signaled via contextual cues, where the descending pitch of the tone denotes the counting-down delay the hungry rat must wait in order to get a reward. The rat can choose to wait out the countdown or skip mid-countdown and move on to the next station, foraging around this "food court" on a limited time budget. (B) Humans similar forage on a computer for natural rewards (short entertaining videos) of varying costs (download lengths) cycling through different "galleries" (video genre). A sequence of "next" button clicks mimic the travel costs rodents experience between restaurants. (C-D) Example data for a single subject's session, demonstrating reliable revealed subjective valuation preferences of willingness to wait in particular restaurant or gallery. Red cross – inflection point or "threshold" of sigmoid fits of choices as a function of cost.

Below, I briefly highlight the discoveries made in this thesis:

- In the second chapter's original research, I demonstrate that mice, like rats, too, are capable of experiencing regret on a novel variant of the Restaurant Row task. I discover hidden costs and hidden utilities associated with regret-related learning that drive the development of distinct decision-making strategies longitudinally that are behaviorally dissociable. I dissociate foraging valuations separate from deliberative valuations that each processes decision conflict differently. I hypothesize that such behaviorally dissociable valuation algorithms arise from separable neural circuits.

- In the fourth chapter's original research, I translate this novel variant of the Restaurant Row task into a novel variant of the Web-Surf Task and demonstrate that mice, rats, and humans both deliberate and forage similarly for rewards. Interestingly, I discover that foraging valuation algorithms but not deliberative valuation algorithms are uniquely susceptible to self-control related cognitive biases preserved across evolution, suggesting that the separable neural circuits from which these valuations arise, too, are preserved.

- In the sixth chapter, I propose a novel unified theoretical perspective describing how advancements in neuroeconomics in light of recent developments in circuit dissection tools can push the field of computational neuropsychiatry forward. I use addiction as a case study and bring together fields of the neurobiology of learning and memory with neuroeconomics of decision-making information processing to reveal how more can be learned and how new questions can be asked to better interrogate disease heterogeneity. Chapters seven and nine perform two critical types of experiments proposed in this new theoretical framework.

- In the seventh chapter's original research, I build off of the first half of the thesis, which described in detail how complex behavioral neuroeconomics can reveal multiple parallel decision-making systems that are characterizable through sophisticated behavioral testing. I find that high-conflict economic decisions (encountering expensive offers for highly desired rewards) operationalizes a sophisticated level of decisions, particularly those related to self-control, not well-modeled in animals until now. I use this economic scenario to operationalize the conflict between "wanting vs. knowing better" – and that mice are able to behaviorally communicate this conflict in separate decision valuations to us. I characterize this conflict in two distinct deliberative and foraging valuation algorithms. I then demonstrate the utility of this approach in a well-established mouse model of addition (based on work in the Thomas lab and others) comparing effects of two different classes of drugs of abuse – psychostimulants (cocaine) vs. opiates (morphine). I test how these sophisticated level of conflict decisions might be uniquely disrupted in separate addiction models in mice – where the vast majority of studies comparing such drugs using simple tests of value report similar behavioral dysfunctions. While I do not directly model addiction per se (compulsive maladaptive drug-seeking behaviors) since mice tested on this task are making decisions in pursuit of natural food rewards, I instead examine how drug-related experiences, which can leave profound lasting plasticity changes in the brain even throughout prolonged abstinence, can uniquely disrupt the ways in which these types of self-control decisions are made. For, little is known about how these types of decisions can be modeled in non-human animals, which brings much needed face validity to approximate the types of complex decisions recovering human addicts struggle with before relapsing. I discover that abstinence from chronic cocaine vs. morphine exposure produce dissociable long-lasting disruptions in separable deliberative and foraging valuation algorithms during these high conflict "wanting vs. knowing better" decisions.

- In the ninth chapter's original research, which summates to the capstone demonstration of multiple, parallel decision-making systems building off of the rest of the earlier chapters of this thesis, I directly interrogate the functional consequences of synaptic remodeling in specific circuits (the glutamatergic infralimbic to accumbens shell circuit) on distinct aspects of neuroeconomic valuations. I use optogenetics in mice trained on this novel variant of the Restaurant Row task in order to interrogate a projection-specific corticostriatal circuit using a plasticity manipulation delivered acutely "off-line," or outside of behavioral testing. I discover that was able to induce long-lasting changes in self-control related foraging valuations independent of deliberative valuations measured within the same trial. Importantly, these changes mimic disrupts seen in morphine-abstinent mice, but not cocaine-abstinent mice. Furthermore, this type of circuit plasticity change has been reported to occur following triggers of relapse that might move an individual into a decision-vulnerable state. I also developed a novel circuit-specific plasticity measurement tool and reveal that individual differences circuit-specific synaptic strength can explain behavioral variability of self-control in foraging valuations but not deliberative valuations.

- In the eleventh chapter, I conclude with a synthesis of the discoveries presented in this thesis. I discuss the implications for an emerging field in mental health clinical practice, Computational Psychiatry, which strives to resolve disease heterogeneity by moving away from pure symptomology as a method for disease diagnosis and treatment toward biomarkers that reflect disruptions in heterogeneous underlying neural computations that give rise to behavioral dysfunction. The collective work presented in this thesis directly speaks to the mission of Computational Psychiatry and helps pave a way forward for the field of translational neuroscience.

Taken together, in this thesis, I use neuroeconomics to reveal novel insights into the complexity and diversity with which the brain processes value, how multiple parallel decision-making systems are conserved across species over evolution, how carefully designed behaviors can more accurately reflect those dissociable underlying neural computations, how circuit-dissection studies can better take these concepts into consideration in refining circuit-specific computational processes, all in service of elucidating fundamentally distinct ways in which the different decisions we make can go wrong in neuropsychiatric illnesses. The work presented in this thesis has direct implications for how novel therapeutic treatments will likely need to be tailored to the individual in a circuit-computation-specific manner, for much of the current state of neuropsychiatric disease treatments, let alone diagnoses, do not speak to the underlying neural computational dysfunction that may be a disease driver in one individual but not another.

– Original Research –

Chapter 2

Mice learn to avoid regret

Abstract

Regret can be defined as the subjective experience of recognizing a mistake has been made and a better alternative could have been selected. The experience of regret is thought to carry negative utility. This typically takes two distinct forms: augmenting immediate post-regret valuations to make up for losses; and, augmenting long-term changes in decision-making strategies to avoid future instances of regret altogether. While the short-term changes in valuation have been studied in human psychology, economics, neuroscience, and even recently in nonhuman-primate and rodent neurophysiology, the latter long-term process has received far less attention, with no reports of regret-avoidance in non-human decision-making paradigms. I trained 31 mice in a novel variant of the Restaurant Row economic decision-making task, in which mice make decisions of whether to spend time from a limited budget to achieve food rewards of varying costs (delays). Importantly, I tested mice longitudinally for 70 consecutive days, on which the task provided their only source of food. Thus, decision strategies were interdependent across both trials and days. I separated principal commitment decisions from secondary re-evaluation decisions across space and time and found evidence for regret-like behaviors following change-of-mind decisions that corrected prior economically disadvantageous choices. Immediately followed change-of-mind events, subsequent decisions appeared to make up for lost effort by altering willingness to wait, decision speed, and pellet consumption speed, consistent with past reports of regret in rodents. As mice were exposed to an increasingly reward-scarce environment, I found they adapted and refined distinct economic decision-making strategies over the time course of weeks to maximize reinforcement rate. However, I also found that even without changes in reinforcement rate, mice transitioned from an early strategy rooted in foraging, to a strategy rooted in deliberation and planning that prevented future regret-inducing change-of-mind episodes from occurring. These data suggest that mice are learning to avoid future regret, independent of and separate from reinforcement rate maximization.

Chapter reprinted with permissions from *PLoS Biology*, modified from:

Sweis BM, Thomas MJ, Redish AD. 2018b. Mice learn to avoid regret. *PLoS Biology* 16(6): e2005853.

Introduction

Regretful experiences comprise those where an individual recognizes he or she could have made a better decision in the past. Humans assert a strong desire to avoid feeling regret (Zeelenberg and Pieters 2007). Regret can have an immediate impact on influencing subsequent valuations, but it can also motivate individuals to learn to avoid future regret-provoking scenarios altogether (Coricelli et al. 2005). Recently, the experience of regret has been demonstrated in nonhuman animals, sharing principal neurophysiological and behavioral correlates of regret with humans (Steiner and Redish 2014; Abe and Lee 2011). However, it remains unclear if nonhuman animals are capable of learning from regret in order to avoid recurring episodes in the future.

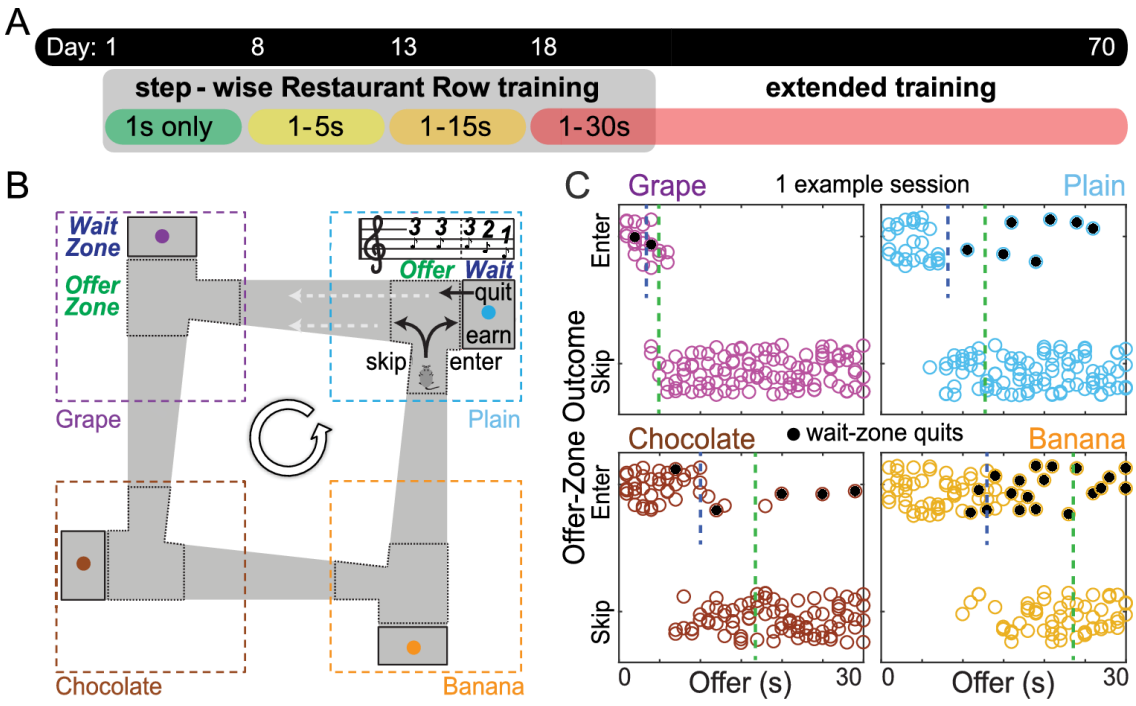
Counterfactual reasoning, or considering what might have been, is a critical tenet of experiencing regret (Epstude and Roese 2008; Byrne 2002). This entails reflecting on potentially better alternatives that could have been selected in place of a recent decision. Thus, owning a sense of choice responsibility and acknowledging error of one's own agency is central to regret. Following the experience of regret, humans often report a change in mood and augment subsequent decisions in an attempt at self-justification or in efforts to make up for their losses (Frydman and Camerer 2016; Coricelli and Rustichini 2010). These immediate effects of regret on behavior describe a phenomenon distinct from the notion that individuals will also learn to take longitudinal measures to avoid future scenarios that may induce regret.

Neuroeconomic decision-making tasks offer a controlled laboratory approach to operationalize and characterize decision-making processes comparable across species (Dickhaut and Rustichini 2009; Kalenscher and van Wingerden 2011; Loewenstein et al. 2008; Rangel et al. 2008). Recently, a study by Steiner and Redish reported the first evidence of regret in rodents tested on a spatial decision-making task (Restaurant Row, Steiner and Redish 2014). In this task, food-restricted rats were trained to spend a limited time budget earning food rewards of varying costs (delays) and demonstrated stable subjective valuation policies of willingness to wait contingent upon cued offer costs. In rare instances when rats disadvantageously violated their decision policies and skipped low cost offers only to discover worse offers on subsequent trials

(e.g., made “economic mistakes”), they looked back at the previous reward site and displayed corrective decisions that made up for lost time. These behaviors coincided with neural representations of retrospective missed opportunities in the orbitofrontal cortex, consistent with human and nonhuman primate reports of counterfactual “might-have-been” representations (Steiner and Redish 2014; Coricelli et al. 2005; Abe and Lee 2011; Camille et al. 2004; Sommer et al. 2009; Steiner and Redish 2012). While these data demonstrate that rats are responsive to the immediate effects of regret, the regret instances were too sparse to determine whether rats also showed long-term consequences of these regret phenomena. Thus, it remains unknown if nonhuman animals are capable of learning from such regret-related experiences, leaving open the question of whether nonhuman animals adopt longitudinal changes in economic decision-making strategies that prevent future instances of regret from occurring in the first place?

In the present study (Figure 2.1), I trained food-restricted mice to traverse a square maze with four feeding sites (restaurants), each with unique spatial cues, each providing a different flavor (Figure 2.1B). On entry into each restaurant, mice were informed of the delay that they would be required to wait to get the food from that restaurant. In this novel variant of the Restaurant Row task, each restaurant contained two distinct zones: an offer zone and a wait zone. Mice were informed of the delay on entry into the offer zone, but delay countdowns did not begin until mice moved into the wait zone. Thus, in the offer zone, mice could either enter the wait zone (to wait out the delay) or skip (to proceed on to the next restaurant). After making an initial enter decision, mice had the opportunity to make a secondary re-evaluative decision to abandon the wait zone (quit) during delay countdowns. Just like rats, mice revealed preferences for different flavors that varied between animals but were stable across days, indicating subjective valuations for each flavor were used to guide motivated behaviors. Varying flavors, as opposed to varying pellet number, allowed us to manipulate reward value without introducing differences in feeding times between restaurants (as time is a limited commodity on this task). Costs were measured as different delays mice would have to wait to earn a food reward on that trial, detracting from their session’s limited 1hr time

Figure 2.1: Novel variant of Restaurant Row adapted for mice tested longitudinally



(A) Experimental timeline. Mice were trained for 70 consecutive days earning their only source of food on this task. Stages of training were broken up into blocks where the range of possible offers began in a reward-rich environment (all offers were always 1s, green epoch) and escalated to increasingly reward-scarce environments (offer ranges of 1-5s, 1-15s, 1-30s). (B) Task schematic. Food-restricted mice were trained to encounter serial offers for flavored rewards in four "restaurants." Restaurant flavor and location were fixed and signaled via contextual cues. Each restaurant contained a separate offer zone and wait zone. Tones sounded in the offer zone; fixed tone pitch indicated delay (randomly selected from that block's offer range) mice would have to wait in the wait zone. Tone pitch descended during delay "countdown" if mice chose to enter the wait zone. Mice could quit the wait zone for the next restaurant during the countdown, terminating the trial. Mice were tested daily for 60 min. (C) Example session (from the 1-30s red epoch) with individual trials plotted as dots. This representative mouse entered low delays and skipped high delays in the offer zone, while sometimes quitting once in the wait zone (black dots). Dashed vertical lines represent calculated offer zone (green) and wait zone (blue) "thresholds" of willingness to budget time. Thresholds were measured from the inflection point of fitting a sigmoid curve to enters vs. skips or earns vs. quits as a function of delay cost.

budget. Delays were randomly selected between a range of offers for each trial. Tones sounded upon restaurant entry, whose pitch indicated offer cost, and descended in pitch stepwise during countdowns once in the wait zone.

Taken together, in this task, mice must make serial judgements in a self-paced manner weighing subjective valuations for different flavors against offer costs, balancing the economic utility of sustaining overall food intake against earning more rewards of a desirable flavor. In doing so, cognitive flexibility and self-control become critical components of decision-making valuation processes in this task assessed in two separate stages of decision conflict (in the offer and wait zones). Importantly, because mice had 1hr to work for their sole source of food for the day, trials on this task were interdependent both within and across days. Therefore, this was an economic task in which time must be budgeted in order to become self-sufficient across days. Here, I tested mice for 70 consecutive days. Thus, the key to strategy development on this task is the learning that takes place across days, for instance, when performance on a given day produces poor yield. Monitoring longitudinal changes in decision-making strategy can provide novel insight into regret-related learning experiences.

Methods

Mice

31-C57BL/J6 male mice, 13-weeks old, were trained in Restaurant Row. Mice were single-housed in a temperature- and humidity-controlled environment with a 12-hr-light/12-hr-dark cycle with water ad libitum. Mice were food restricted to a maximum of 85% free feeding body weight and trained to earn their entire day's food ration during their 1-hr Restaurant Row session. Experiments were approved by the University of Minnesota Institutional Animal Care and Use Committee. Mice were tested at the same time every day in a dim-lit room, were weighed before and after every testing session, and were fed a small post-session ration in a separate waiting-chamber on rare occasions to prevent extremely low weights according to IACUC standards (not <85% free-feeding weights). Previous studies using this task yielded reliable behavioral findings with minimal variability in at least sample sizes of n=4 rodents.(Steiner and Redish 2014)

Pellet training

Mice underwent 1 week of pellet training prior to the start of being introduced to the Restaurant Row maze. During this period, mice were taken off of regular rodent chow and introduced to a single daily serving of BioServ full nutrition 20 mg dustless precision pellets in excess (5g). This serving consisted of a mixture of chocolate-, banana-, grape-, and plain-flavored pellets. Next, mice (hungry, before being fed their daily ration) were introduced to the Restaurant Row maze 1 day prior to the start of training and were allowed to roam freely for 15 min to explore, get comfortable with the maze, and familiarize themselves with the feeding sites. Restaurants were marked with unique spatial cues. Feeding bowls in each restaurant were filled with excess food on this introduction day.

Restaurant Row training

Task training was broken into 4 stages. Each daily session lasted for 1hr. At test start, one restaurant was randomly selected to be the starting restaurant where an offer was made if mice entered that restaurant's T-shaped offer-zone from the appropriate direction in a counter-clockwise manner. During the first stage (day 1-7), mice were trained for 1 week being given only 1s offers. Brief low pitch tones (4000Hz, 500ms) sounded upon entry into the offer-zone and repeated every second until mice skipped or until mice entered the wait-zone after which a pellet was dispensed. To discourage mice from leaving earned pellets uneaten, motorized feeding bowls cleared any uneaten pellets upon restaurant exit. Left over pellets were counted after each session and mice quickly learned to not leave the reward site without consuming earned pellets. The next restaurant in the counter-clockwise sequence was always and only the next available restaurant where an offer could be made such that mice learned to run laps encountering offers across all four restaurants in a fixed order serially in a single lap. During the second stage (day 8-12), mice were given offers that ranged from 1s to 5s (4000Hz to 5548Hz, in 387Hz steps) for 5 days. Offers were pseudo-randomly selected such that all 5 offer lengths were encountered in 5 consecutive trials before being re-shuffled, selected independently between restaurants. Again, offer tones repeated every second in the offer-zone indefinitely until either a skip or enter decision was made. In this stage and subsequent stages, in the wait-zone, 500ms tones descended in pitch every second by 387Hz steps counting down to pellet delivery. If the wait-zone was

exited at any point during the countdown, the tone ceased and the trial ended, forcing mice to proceed to the next restaurant. Stage 3 (day 13-17) consisted of offers from 1s to 15s (4000Hz to 9418Hz) for another 5 days. Stage 4 (day 18-70) offers ranged from 1s to 30s (4000Hz to 15223Hz) and lasted until mice showed stable economic behaviors. I used 4 Audiotek tweeters positioned next to each restaurant powered by Lepy amplifiers to play local tones at 70dB in each restaurant. I recorded speaker quality to verify frequency playback fidelity. I used Med Associates 20mg feeder pellet dispensers and 3D-printed feeding bowl receptacles fashioned with mini-servos to control automated clearance of uneaten pellets. Animal tracking, task programming, and maze operation was powered by AnyMaze (Stoelting). Mice were tested at the same time every day in a dim-lit room, were weighed before and after every testing session, and were fed a small post-session ration in a separate waiting chamber on rare occasions as needed to prevent extremely low weights according to IACUC standards (not <85% free-feeding weights).

Statistical analysis

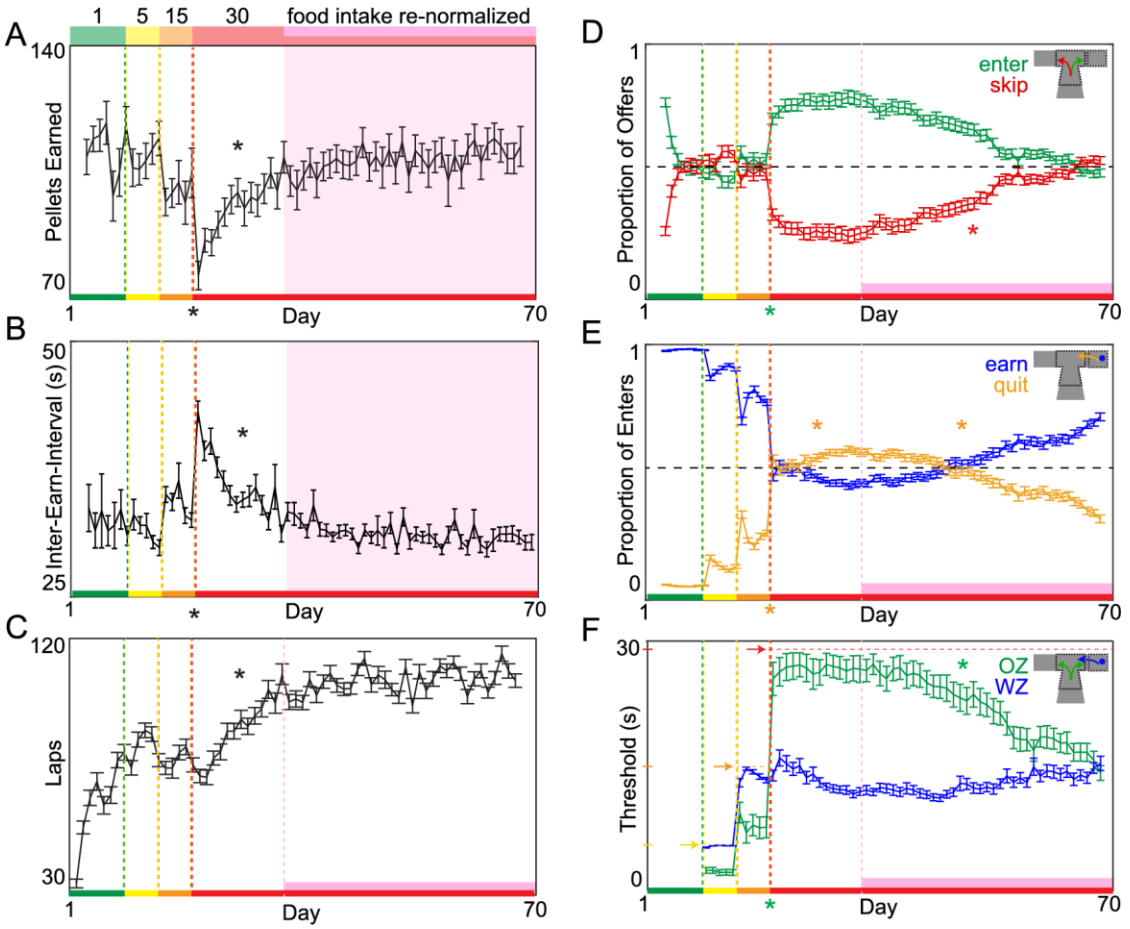
All data were processed in Matlab and statistical analyses were carried out using JMP Pro 13 Statistical Discovery software package from SAS. All data are expressed as mean +/- 1 standard error. Sample size is included in each figure. No data were lost to outliers. Offer zone thresholds were calculated by fitting a sigmoid function to offer zone choice outcome (skip vs. enter) as a function offer length for all trials in a single restaurant for a single session and measuring the inflection point. Wait zone thresholds were calculated by fitting a sigmoid function to wait zone choice outcomes (quit vs. earn) as a function of offer length for all entered trials in a single restaurant for a single session. For dynamic analyses that depend on thresholds, analyses at each timepoint used that timepoint's threshold information. Statistical significance was assessed using student's t tests, one-way, two-way, and repeated measures ANOVAs, using mouse as a random effect in a mixed model, with post-hoc Tukey t tests correcting for multiple comparisons. Significance testing of immediate changes at block transitions were tested using a repeated measures ANOVA between 1 day pre and 1 day post transition. These are indicated by significance annotations below the x-axis on relevant figures. Significance testing of gradual changes within block were tested using a repeated measures ANOVA across all days within a given block or epoch. These are indicated by significance annotations within the plot either

directly above or below the data centered within the epoch of interest. If interactions between conditioned were tested (e.g., x rank), these are reflected by multiple significance annotations either below the x-axis or within the plot, respectively. The period of renormalization was estimated based on animal self-driven performance improvements in the 1-30s block and not imposed on the animals by experimenters nor the protocol design. Re-normalization was characterized by identifying the number of days in the 1-30s block after which total pellet earnings and reinforcement rate reliably stabilized (within a sliding 5-day window) and was no different from performance in relatively reward-rich environments collapsing across the first three training blocks. This was estimated to be approximately by day 30 of the experiment.

Results

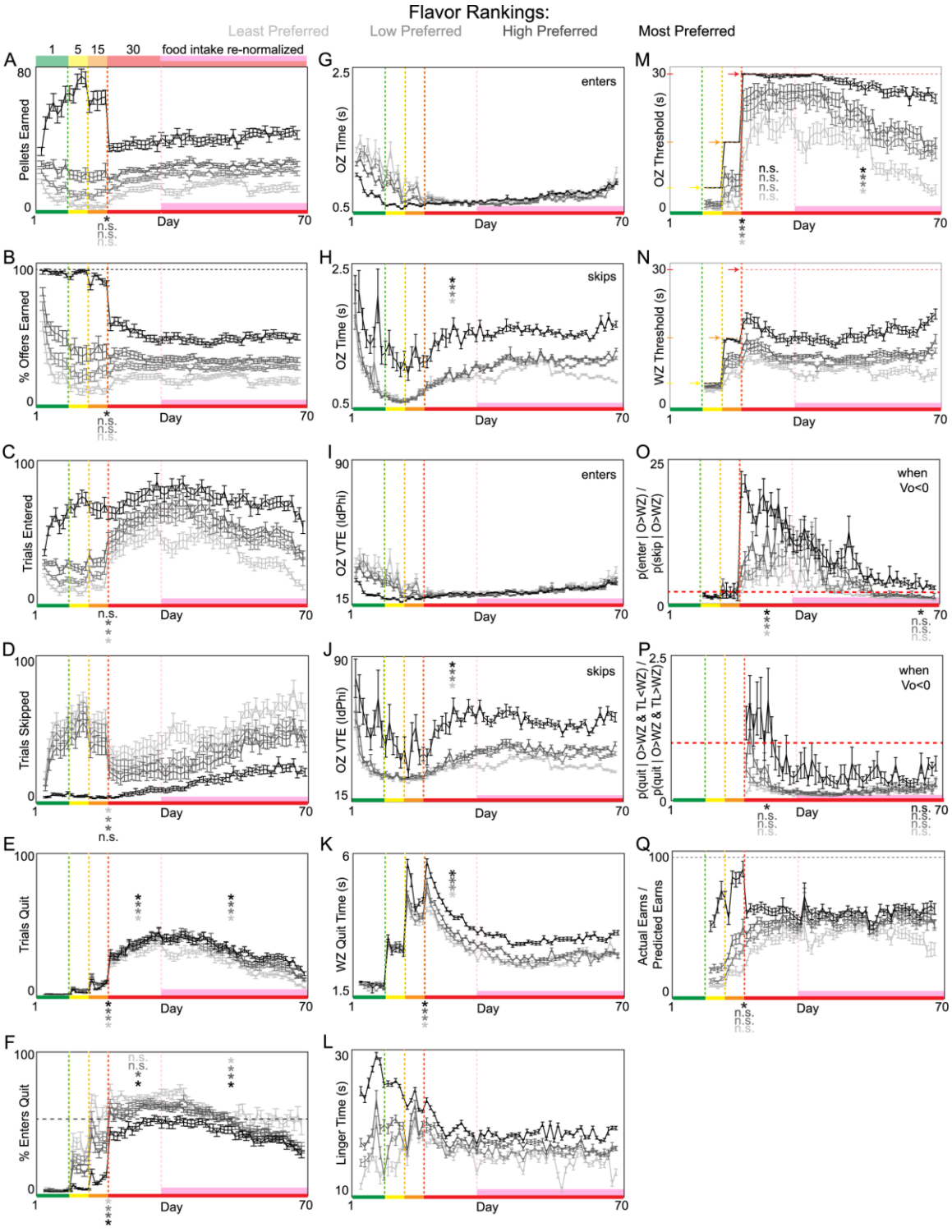
How mice were trained on the Restaurant Row task allowed us to characterize the development of and changes in economic decision-making strategies. Mice progressed from a reward-rich to a reward-scarce environment in blocks of stages of training across days (Figure 2.1A). Each block was defined by the range of possible costs that could be encountered when offers were randomly selected on the start of each trial upon entry into each restaurant's offer zone. The first block (green epoch) spanned 7 days where all offers were always 1s (Figure 2.1A). During this time mice quickly learned the structure of the task, becoming self-sufficient and stabilizing number of pellets earned (Figure 2.2A), reinforcement rate (Figure 2.2B), and number of laps run (Figure 2.2C). Flavors were ranked from least preferred to most preferred based on total pellet earnings in each restaurant at the end of each session (Figure 2.3A). During this first block, mice rapidly developed stable flavor preferences and learned to skip offers for less-preferred flavors and enter offers for more-preferred flavors, entering vs. skipping at roughly equal rates overall, while rarely quitting (Figure 2.2D-E, Figure 2.3A-E). The second block (yellow epoch) spanned 5 days where offers could range

Figure 2.2: Changes in economic decisions in an increasingly reward-scarce environment



(A-B) Primary dependent variables: total earned food intake (A) and reinforcement rate (B, measured as average time between earnings). Transition to the 1-30s block caused a significant decrease in food intake and reinforcement rate. By approximately day 32, food intake and reinforcement rate re-normalized back to stable baseline levels compared to previous testing in reward-rich environments. The epoch marked in pink defines this re-normalization to baseline and is used throughout the remaining longitudinal plots. (C) Number of self-paced laps run (serially encountering an offer in each of the four restaurants). (D) Proportion of total offers entered vs. skipped. Horizontal dashed line represents 0.5 level. (E) Proportion of total enters earned vs. quit. Horizontal dashed line represents 0.5 level. (F) Economic decision thresholds: offer zone and wait zone choice outcomes as a function of cost. Horizontal dashed lines represent the maximum possible threshold in each block. Data are presented as the cohort's (N=31) daily means (±1SE) across the entire experiment. Color code on the x-axis reflects the stages of training (offer cost ranges denoted on the top of panel A). Vertical dashed lines (except pink) represent offer block transitions. * on the x-axis indicates immediate significant behavioral change at the block transition, otherwise * indicates gradual significant changes within the 1-30s block during either the 2wk adaptation period or pink epoch.

Figure 2.3: Subjective flavor preferences and longitudinal economic decision processes



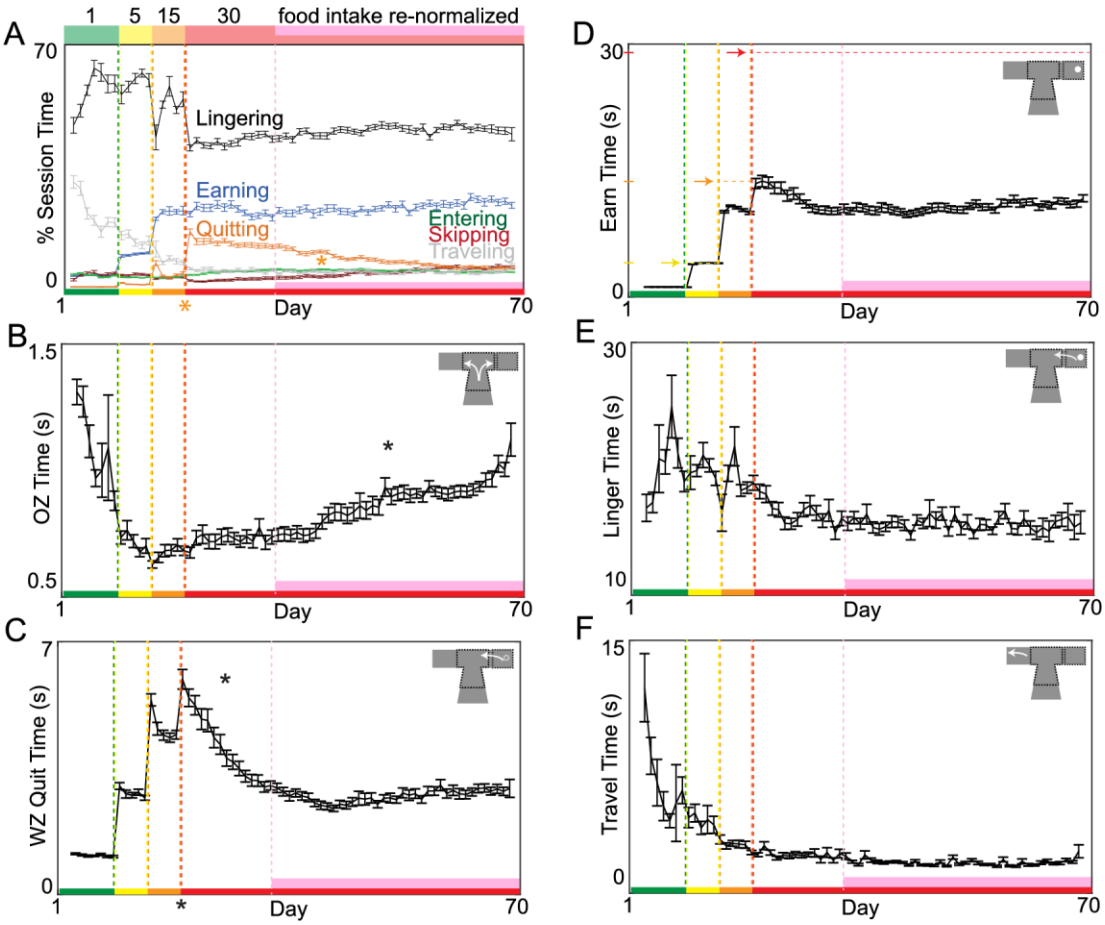
Flavors were ranked from least preferred to most preferred based on total pellet earnings in each restaurant at the end of each session. (A) Pellets earned in each restaurant show early development of flavor preferences that persist throughout the entire experiment. (B) Percentage of offers entered. Horizontal dashed line indicates 100%. (C-E) Total number of trials entered (C), skipped (D), and quit (E). (F) Percentage of entered offers quit. Horizontal dashed line indicates 50%. (G-J) Offer zone behaviors for enter (G, time; I, VTE) and skip (H, time; J, VTE) decisions. (K) Time spent in the wait zone during tone countdown before quitting. (L) Time spent in the wait zone consuming an earned food pellet and lingering near the reward site before advancing to the next trial. (M-N) Offer zone (M) and wait zone (N) thresholds. Horizontal dashed lines represent the maximum possible threshold in each block. (O) Offer zone inefficiency ratio. Offer value (VO) = wait zone threshold – offer cost. Probability of entering negatively valued offers relative to the probability of skipping negatively valued offers. Horizontal dashed line indicates equivalent 1:1 ratio of entering vs. skipping negatively valued offers. (P) Wait zone inefficiency ratio. Value of time left in countdown at the moment of quitting (VL) = wait zone threshold – countdown time left. Probability of quitting negatively valued offers when VL was positive relative to when VL was still negative. Horizontal dashed line indicates equivalent 1:1 ratio of quitting inefficiently vs. efficiently. (Q) Reward-earning optimality. Proportion of pellets mice actually earned in each restaurant relative to model-estimated maximal predicted earnings. Horizontal dashed line indicates 100% optimal earnings. Data are presented as the cohort's (N=31) daily means (±1SE) across the entire experiment. Color code on the x-axis reflects the stages of training (offer cost ranges denoted on the top of panel A). Vertical dashed lines (except pink) represent offer block transitions. * on the x-axis indicates immediate significant behavioral change at the block transition. * on the x-axis in (O-P) indicates significantly inefficient decisions (above the 1:1 efficiency ratio line). Otherwise, * indicates gradual significant changes within the 1-30s block during either the 2wk adaptation period or pink epoch. Not significant (n.s.).

between 1-5s. The third block (orange epoch) spanned 5 days where offers could range between 1-15s. Lastly, the fourth and final block (red epoch, beginning on day 18) lasted until the end of this experiment (day 70) where offers could range between 1-30s. Note that because the mice had a limited 1hr time budget to get all of their food for the day, these changes in offer distributions produced increasingly reward-scarce environments that required more complex strategies to maximize, or merely maintain, rate of reward.

Somewhat surprisingly, beginning on day 1 of training, I found that mice, after earning and consuming food pellets on the Restaurant Row task, often lingered at the reward site before advancing to the next trial at the next restaurant. Interestingly, rodents typically spent >50% of the entire 1 hr testing session engaging in this lingering behavior (Figure 2.4A, t=14.66, p<0.0001). The decision to linger near the reward site rather than leave may represent a strong conditioned-place-preference-like effect associated with each restaurant's unique spatial context.(Clark et al. 2012) This decision depletes an animal's limited time budget and comes with the cost of impeding earning a subsequent reward at the next restaurant since self-paced trials are interdependent. Thus, decisions to linger indeed reflect a valuation process.

The value of a reward can be assessed in multiple ways. Often, this is measured in an instrumental manner, as how willing a subject is to take a reward (e.g., measured in amount of resources spent, effort expended, number of reward opportunities seized, or behavioral invigoration while retrieving rewards). These behaviors are sometimes referred to as reward-seeking, reward-taking, or "wanting" valuations (Berridge 1996). These can be separated from more general hedonic valuations, often called "liking," that are usually measured post-consumption (Berridge 1996). On this task, post-consumption lingering behavior occur *after* a reward has been earned, where no overt reward is being sought out. Thus, these valuations appear distinct from reward-taking or "wanting" valuations that may occur during offer zone and wait zone decisions *before* a reward is earned. Therefore, lingering behavior may reflect a distinct, although related, valuation embedded in contextual Pavlovian associations (Berridge 1996; Clark et al. 2012).

Figure 2.4: Allocation of a limited time budget among separable decision processes



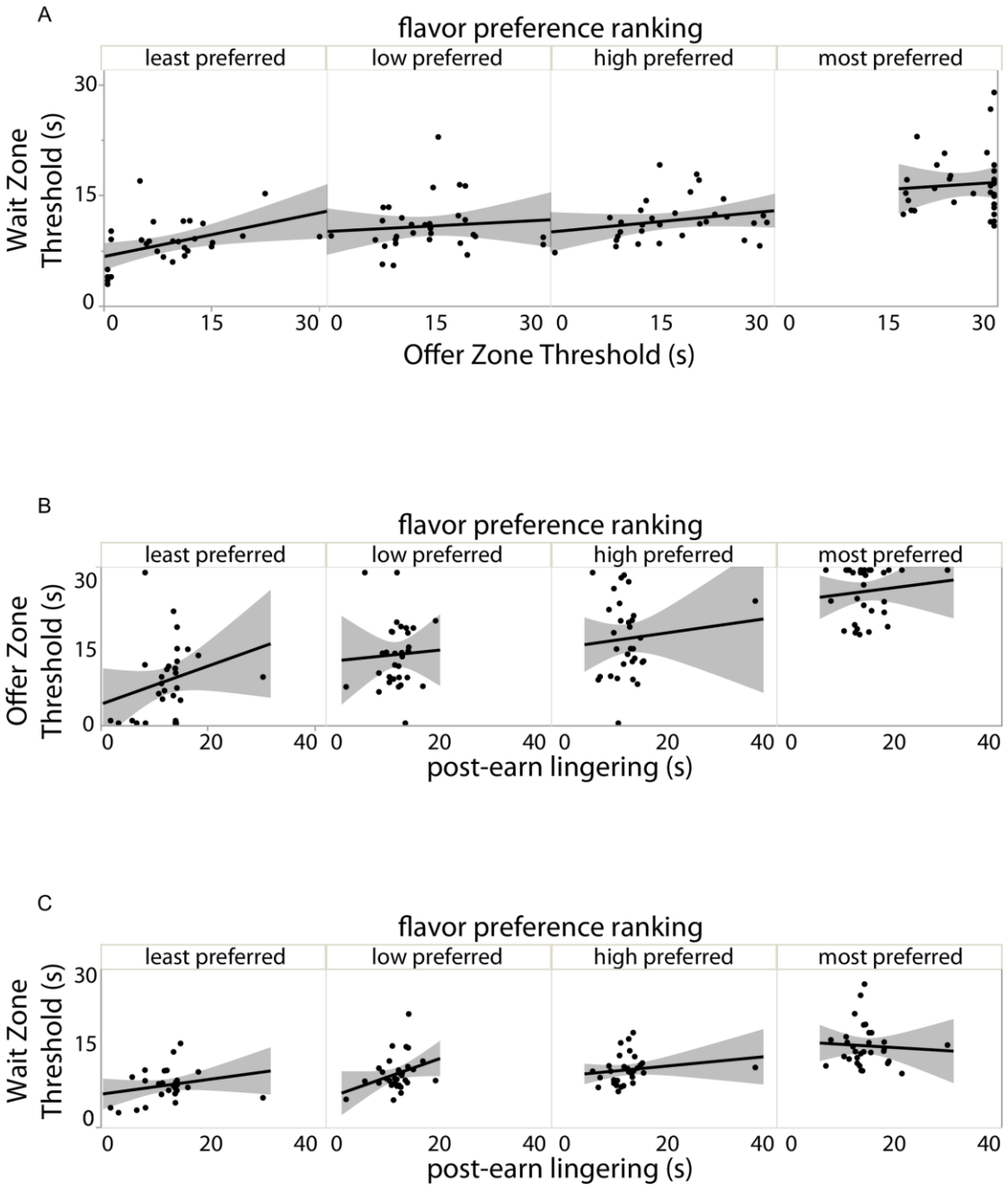
(A) Cumulative time spent engaged in various separable behaviors and decision processes calculated as percent of the total 1hr daily session's time budget. (B) Average time in the offer zone from offer onset upon restaurant entry until either a skip or enter decision was made. (C) Average time in the wait zone from countdown onset until a quit decision was made. (D) Average time in the wait zone from countdown onset until a pellet was earned. (E) Average time near the reward site from pellet delivery until mice exited the wait zone and entered the hallway advancing to the next restaurant. (F) Average time spent traveling in the hallway between restaurants between trials (from either a skip, quit, or post-earn leave decision until the next trial's offer onset upon subsequent restaurant entry). Data are presented as the cohort's (N=31) daily means (±1SE) across the entire experiment. Color code on the x-axis reflects the stages of training (offer cost ranges denoted on the top of panel A). Vertical dashed lines (except pink) represent block transitions. * on the x-axis indicates immediate significant behavioral change at the block transition, otherwise * indicates gradual significant changes within the 1-30s block during either the 2wk adaptation period or pink epoch.

Evidence for this is garnered by the fact that mice lingered longer after earning a reward in higher preferred restaurants (Figure 2.3L, F=365.73 p<0.0001). Thus, the context of a specific restaurant may be able to communicate such Pavlovian associations. This matches the higher frequency of offer zone enter decisions and wait zone earn decisions made in higher preferred restaurants. As I will demonstrate later, context-associated Pavlovian valuations that may be promoting lingering may similarly carry added weight in promoting offer zone decisions to enter and wait zone decisions earn. So, while these three decision processes – enter vs. skip in the offer zone, earn vs. quit in the wait zone, and linger vs. leave the reward site – may be in register with the ordinal ranking of revealed subjective preferences of each restaurant, they remain fundamentally distinct behavioral computations (Figure 2.5). The importance of these distinctions will become apparent when separable learning processes interact with each of these behavioral computations uniquely.

Upon transitioning to the 1-30s offer block, mice suffered a large drop in total number of pellets earned (Figure 2.2A, repeated measures ANOVA, F=9.46, p<0.01) and reinforcement rate (increase in time between earnings, Figure 2.2B, F=253.93, p<0.0001). With this came a number of changes in decision-making behaviors that took place immediately, on an intermediate timescale, and on a delayed long-term timescale. Decreases in food intake and reinforcement rate were driven by an immediate significant increase in proportion of total offers entered (Figure 2.2D, F=56.10, p<0.0001) coupled with a significant increase in proportion of entered offers quit (Figure 2.2E, F=472.88, p<0.0001) as mice experienced long delays in the wait zone for the first time. This suggests that mice were apt to accept expensive offers in the offer zone even though they did not actually earn those offers in the wait zone (Figure 2.6C,G,I-J). This also suggests that choosing to enter versus skip in the offer zone and choosing to opt out of waiting in the wait zone may access separate valuation algorithms in addition to being physically different action-selection processes.

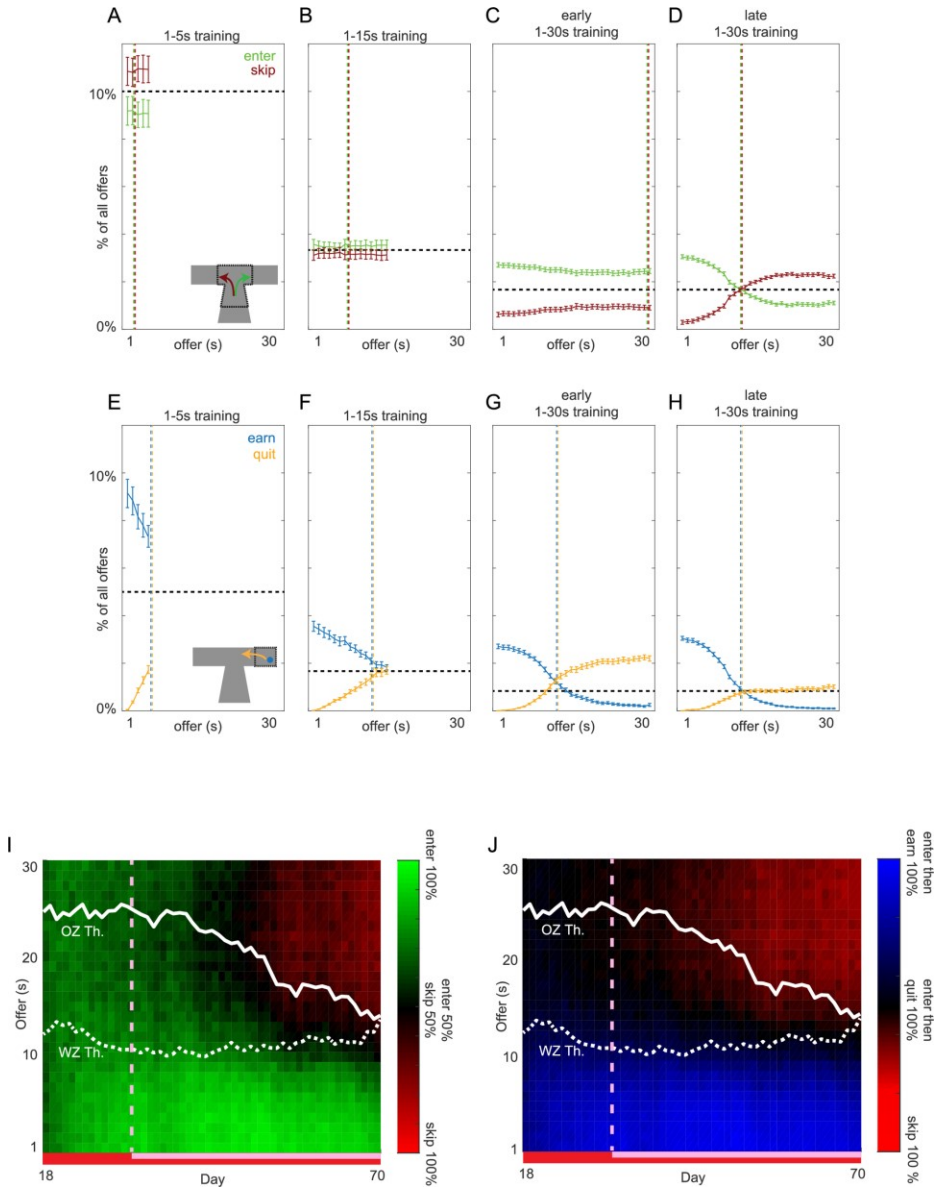
I quantified this disparity in economic valuations by calculating separate “thresholds” of willingness-to-enter in the offer zone and willingness-to-wait in the wait zone as a function of offer cost. Each of these thresholds

Figure 2.5: Independent valuations across offer-zone, wait-zone, and post-earn lingering behaviors



(A-C) Outside of the ordinal rankings of subjective flavor preferences, no relationships were observed between offer-zone and wait-zone thresholds (A), offer-zone thresholds and post-earn lingering time (B), or wait-zone thresholds and post-earn lingering time. All correlations, correcting for multiple comparisons, resulted in non-significance, $P>0.05$.

Figure 2.6: Decision outcomes as function of offer costs across training



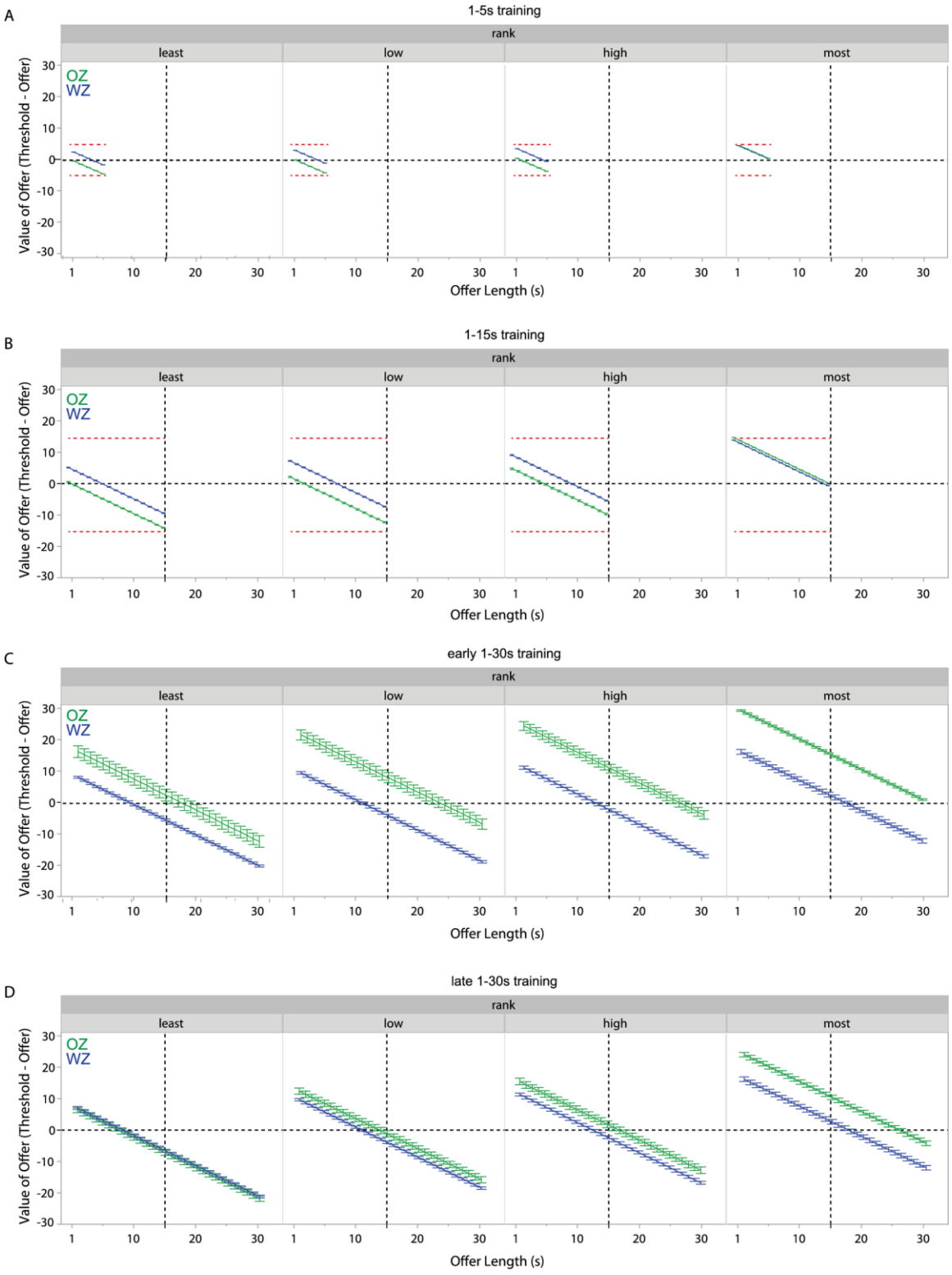
Choice probability to enter vs. skip in the offer zone (A-D) or earn vs. quit in the wait zone (E-H) relative to all offers (normalized to all session trials) as a function of cost during 1-5s training (A,E), 1-15s training (B,F), early 1-30s training (C,G, first 5 days), and late 1-30s training (D,H, last 5 days). Vertical dashed lines indicate average threshold. Wait zone thresholds remained relatively stable across 1-15s and 1-30s offer blocks. Offer zone thresholds became in register with wait zone thresholds by the end of the 1-30s training. Horizontal dashed lines indicate choice probability if decisions were made at random. Data are presented as the cohort's (N=31) means (±1SE). (I-J) Offer zone outcome (I) and trial-end outcome (J) probabilities as a function of offer cost over the 1-30s training block (red epoch). All subjects pooled together for visualization purposes. Solid white line represents cohort's overall average offer zone threshold. Dashed white line represents cohort's overall average wait zone threshold. Pink line represents onset of food intake and reinforcement rate re-normalization after 2wks of adaptation following the transition to 1-30s offers (pink epoch spans days 32-70).

can capture distinct subjective economic valuations of a given reward offer (Figure 2.7, Figure 2.8). Following the 1-30s transition, offer zone thresholds significantly increased (maxed out at ~30s) and became significantly higher than wait zone thresholds (Figure 2.2F, offer zone change: F=151.65, p<0.0001; offer zone vs. wait zone: F=59.85, p<0.0001). Furthermore, I found that these immediate behavioral changes were more robust in higher preferred restaurants, suggesting asymmetries in sub-optimal decision-making strategies upon transition from a reward-rich to a reward-scarce environment were dependent on differences in subjective valuation algorithms (Figure 2.3A-F).

Because performance on this task served as the only source of food for these mice, decision-making policies that might have been sufficient in reward-rich environments must change when they are no longer sufficient in reward-scarce environments. I found that mice demonstrated behavioral adaptations over the 2 weeks following the transition to the 1-30s offer range so that by approximately day 32, they had effectively restored overall food intake (Figure 2.2A, change across 2wks: F=355.21, p<0.0001; post-2wks compared to baseline: F=0.80, p=0.37) and reinforcement rates (Figure 2.2B, change across 2wks: F=183.68, p<0.0001; post-2wks compared to baseline: F=0.24, p=0.63) to baseline levels similar to what was observed in a reward-rich environment (Figure 2.2A-B). Note that the restored reinforcement rates renormalization indicated by the pink epoch in Figure 2.2 was not imposed by the experimenters, but was due to self-paced changes in the behavior of the mice under unchanged experimental rules (red epoch, 1-30s offers).

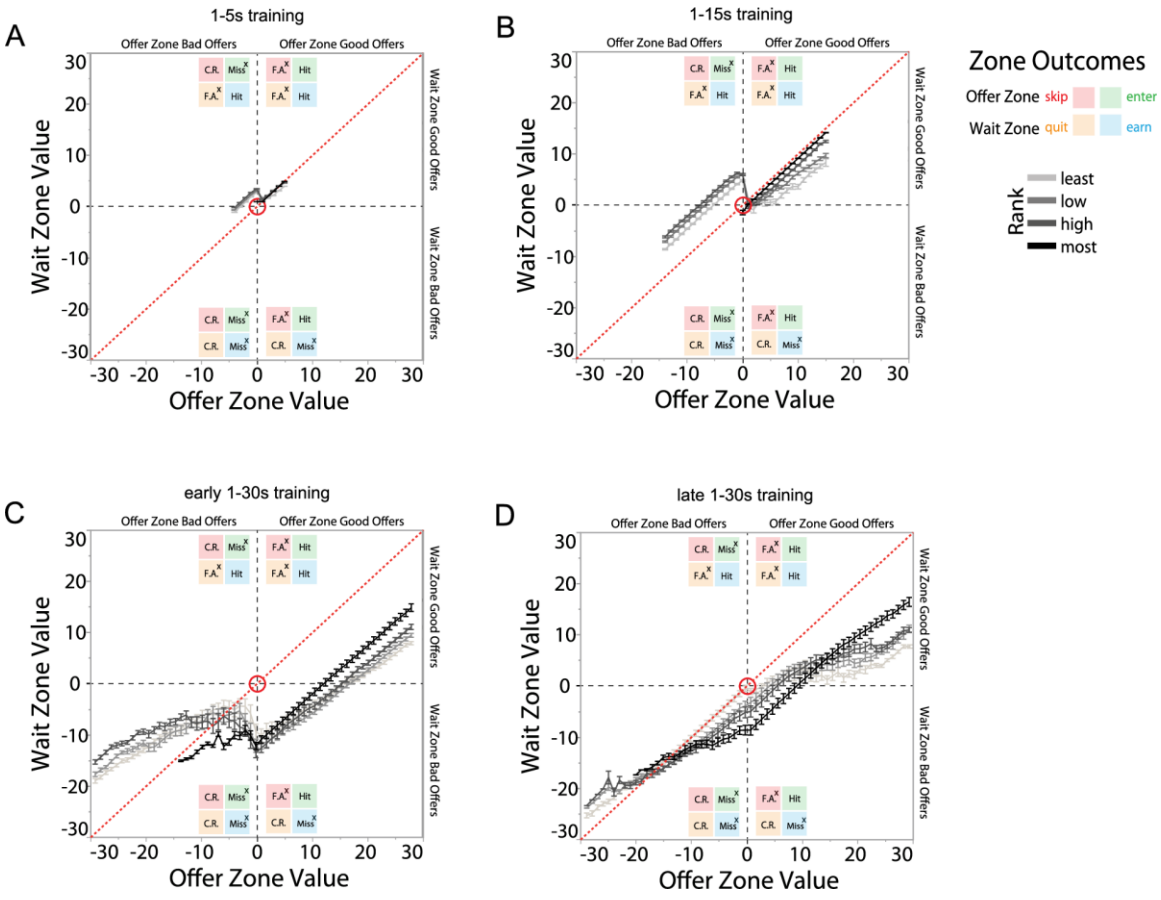
Mice accomplished this by running more laps to compensate for food loss (Figure 2.2C, F=221.61, p<0.0001) without altering economic decision-making policies. That is, I observed no changes in thresholds during this 2wk period (Figure 2.2F, F=2.57, p=0.11). By entering the majority of offers indiscriminately with respect to cost (Figure 2.2D, proportion trials entered > 0.5: t=31.22, p<0.0001, Figure 2.3C), mice found themselves sampling more offers in the wait zone they were unwilling to completely wait for, leading to an increase in quitting (Figure 2.2E, F=55.37, p<0.0001, Figure 2.6G).

Figure 2.7: Offer value defined by zone thresholds



The relative value of an offer on a given trial can be determined by normalizing the offer to the that restaurant's daily threshold. That is, Value = Threshold – Offer. The value of an offer in the offer zone vs. the value of an offer in the wait zone can be calculated by using the offer zone and wait zone threshold in the above equation, respectively. Stages of training are depicted across (A-D), split by the subjective preference ranking of each restaurant. The vertical dashed black lines indicate 15s offers and the horizontal dashed black lines indicate 0 value (where the offer of a given trial is equal either to the offer zone or wait zone thresholds). The red bounds in (A) and (B) indicate the bounds of possible value scores during those training blocks. (D) Early-1-30s training, offer value computed relative to offer-zone thresholds were always higher than relative to wait-zone thresholds, in all restaurants ($P_{d18\text{-}28}<0.0001$, ANOVAs). (B) Late-1-30s training, offer-zone-threshold-derived value was higher than wait-zone-threshold-derived value more so in higher preferred restaurants but not in least preferred restaurants ($P_{d60\text{-}70,most}<0.0001$, $P_{d60\text{-}70,high}<0.001$, $P_{d60\text{-}70,low}<0.05$, $P_{d60\text{-}70,least}>0.05$, ANOVAs).

Figure 2.8: Offer zone value vs. wait zone value definitions of good and bad deals



These plots illustrate the relationship between offer-zone-threshold-derived value and wait-zone-threshold-derived value, particularly highlighting the disparity between indifference points (zero value). Furthermore, these plots illustrate four different types of economic scenarios in each of the four quadrants and allow for the operationalization of relatively "good" vs. "bad" deals in the offer zone and in the wait zone that, according to an individual's typical economic behavior, indicate what decisions one "ought" to make given that scenario. Thus, each quadrant has 4 small colored squares that indicate whether or not a given decision outcome in that economic scenario (color) is either a hit, miss, falsa alarm, or correct rejection, to use Signal Detection Theory terminology (for use in Fig. 2.14). X's in these colored boxes denote if that action selected in that economic scenario is an "error." The diagonal red dashed line as well as the red circle at the origin indicate ideal circumstances where both offer zone valuations and wait zone valuations are in register with each other. Note the changes from (A-B) to (C), upon transition to a reward scarce (1-30s) environment, offer value in the offer zone deviates from the red line in all restaurants. The easiest visualization of this is noticing the downward deviation from the red circle at the origin (in A-B, offer zone and wait zone valuations are in register at zero values; in C, offer zone values are greater than wait zone values in all restaurants). After extended training in (D), lesser preferred restaurants migrate back upward toward the origin more so than higher preferred restaurants, where least preferred restaurants indifference points are back in register in the offer zone and wait zone. Because wait zone thresholds determine the distributions of offer costs where a reward was actually earned or not, offer value calculated using wait zone thresholds is a more appropriate metric of offer value. Error-bars ±1 SEM.

Becoming accustomed to selecting default responses in a reward-rich environment where offer cost is irrelevant can become problematic when suddenly transitioning into a novel reward-scarce environment. The significant decrease in food intake and reinforcement rate observed upon transition to the 1-30s offer block (Figure 2.2A-B) can be entirely explained by a loss of earnings in solely the most preferred restaurant driven by adhering to pre-cost-change decision policies (Figure 2.3A, F=20.75 p<0.0001).

Economic theory of demand elasticity posits that purchase behaviors for luxury items respond most robustly to increases in market prices such that individuals become less capable or willing to continue purchasing such goods if fixed with the same income (van Wingerden et al. 2015). That is, more preferred goods are usually the more elastic ones. Conversely, highly inelastic goods are those that do not respond as much to price changes and often reflect either essential or lesser preferred goods. Using pellets earned in each restaurant as our primary dependent variable, mice were unable to continue to afford earning most preferred pellets in a reward-scarce environment in the same amount that they were previously accustomed to in reward-rich environments (Figure 2.3A). Therefore, most preferred flavors were most elastic in that they suffered the largest change / drop in pellet earnings. However, the reasoning for this appears to be due to decision policies that did not change. From this perspective, pre-cost-change decision policies that remain unchanged, or even stubborn and resist change, in an increasingly reward-scarce environment can reveal interesting complexities of demand elasticity theory and decision processes.

Our data reveal demand elasticity asymmetries in mouse economic strategies as a function of subjective flavor preferences following an unexpected price change when monitored longitudinally. Demand elasticity is well-studied in human microeconomics, however, animal neuroeconomics has only recently started to explore the decision-making phenomena underlying demand elasticity. A recent neuroeconomics study in rats on a different task demonstrated similar aspects of demand theory of elasticity where budget constraints interact with subjective flavor preferences when reward prices escalated depending on whether or not an individual budget is compensated with changing costs (van Wingerden et al. 2015). Here, in the present study, because mice were allotted the same 1 hr time budget that remained fixed across the transition from reward-

rich to reward-scarce environments (and thus went uncompensated) and because mice were tasked with earning their only source of food for the day interdependent across days, mice consequently suffered an initial loss in food intake. This was largely due to adhering to previously-learned default decision policies in each restaurant that become insufficient in a reward-scarce environment and thus was apt to produce poor yield. As a result, mice were pressured to augment decision strategies over subsequent days / weeks. A major advantage in the present study is the longitudinal nature of this neuroeconomic task. To date, no animal neuroeconomic studies have reported a longitudinal account of how subjective value-driven demand elasticity theory manifests over (a) pre-price-change behaviors, (b) immediate negative consequences of unexpected price change, and (c) intermediate-to-long-term strategy changes that learn to adapt in a self-paced manner and learn to work with a fixed, uncompensated time budget.

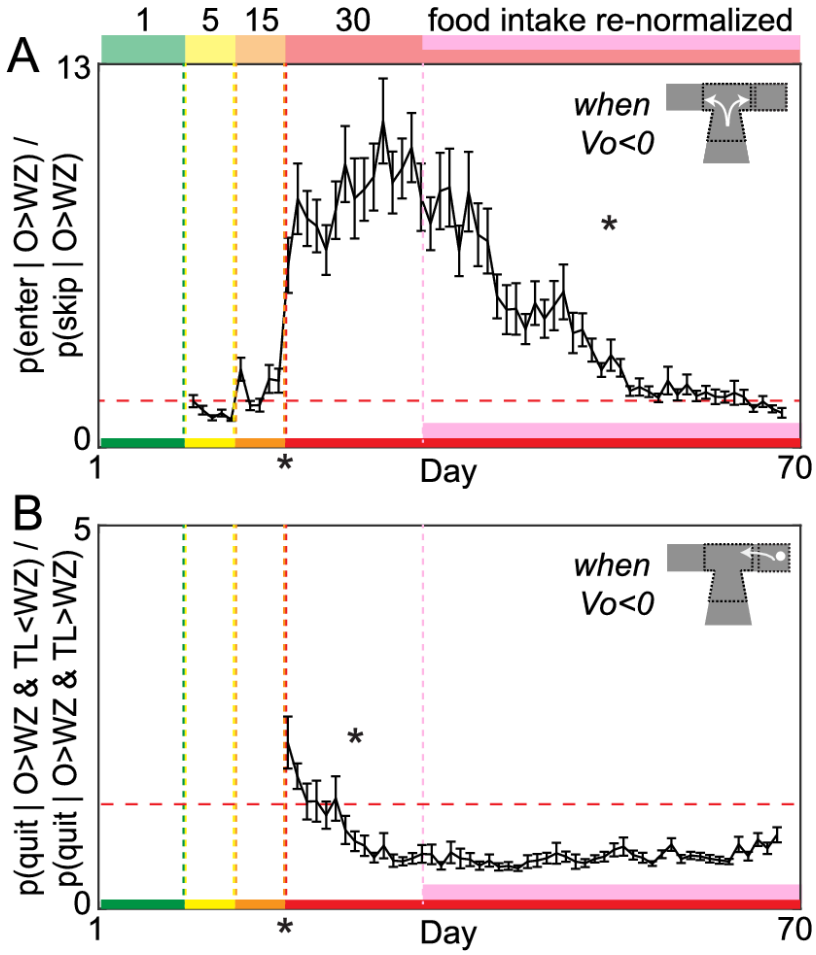
Immediately following the transition to 1-30s offers, the significant increase in quitting – that is, investing a greater portion of a limited time budget waiting for rewards that are ultimately abandoned – appears, at face value, to be a wasteful decision-making strategy. Yet mice were able to restore food intake and reinforcement rates using this strategy. I characterized how mice allocated their limited time budget and quantified time spent among various separable behaviors that made up the total 1hr session (Figure 2.4). I first calculated the percent of total budget engaged in making offer zone decisions to skip vs. enter, wait zone decisions to quit vs. earn, post-earn consumption behaviors, and travel time between restaurants (Figure 2.4A). I also calculated the average time spent engaged in a single bout of each decision process (Figure 2.4B-F). The percent of total session time allocated to quit events (Figure 2.4A, F=306.72, p<0.0001), as well as average time spent waiting before quitting (Figure 2.4C, F=44.21, p<0.0001) significantly increased immediately following the transition to 1-30s offers. Thus, time spent waiting in the wait zone before engaging in change-of-mind behaviors drove the immediate decrease in reinforcement rates and overall loss of food intake. Note that this waiting and then quitting behavior entails investing time that provided no reward. Over the subsequent 2 weeks, time spent waiting before quitting significantly decreased as mice restored food intake and reinforcement rates (Figure 2.4C, F=781.55, p<0.0001). This suggests that mice learned to quit more efficiently in the wait zone.

I calculated economic efficiency of wait zone quits (Figure 2.9B) by measuring how much time was remaining in the countdown at the moment of quitting relative to an individual's wait zone threshold. Over these 2 weeks, mice learned to quit in a more economically advantageous manner before excess time was invested. That is, mice learned to quit while the time remaining in the countdown was still above wait zone thresholds (Figure 2.9B, F=64.00, p<0.0001, Figure 2.3P, Figure 2.10), avoiding quitting at a timepoint when it would have been advantageous to otherwise finish waiting. This suggests that wait zone quit re-evaluations were corrective actions that opposed erroneous principal valuations in the offer zone.

Interestingly, mice struggled to learn to quit efficiently in more preferred restaurants, reflecting a reluctance to apply adaptive opt-out foraging strategies in situations with high subjective valuation biases (Figure 2.3K,P). I call this a foraging process because it reflects a classic, well-studied "abandon current work" decision "in search of something better" often observed in non-human animals in the wild during naturalistic food-seeking behaviors (Stephens and Krebs 1986). Despite increasing change-of-mind efficiency, because the frequency of quit events increased along this 2-week time course, the fraction of the session budget allocated to quit events remained significantly elevated compared to baseline (Figure 2.4A, F=105.90, p<0.0001).

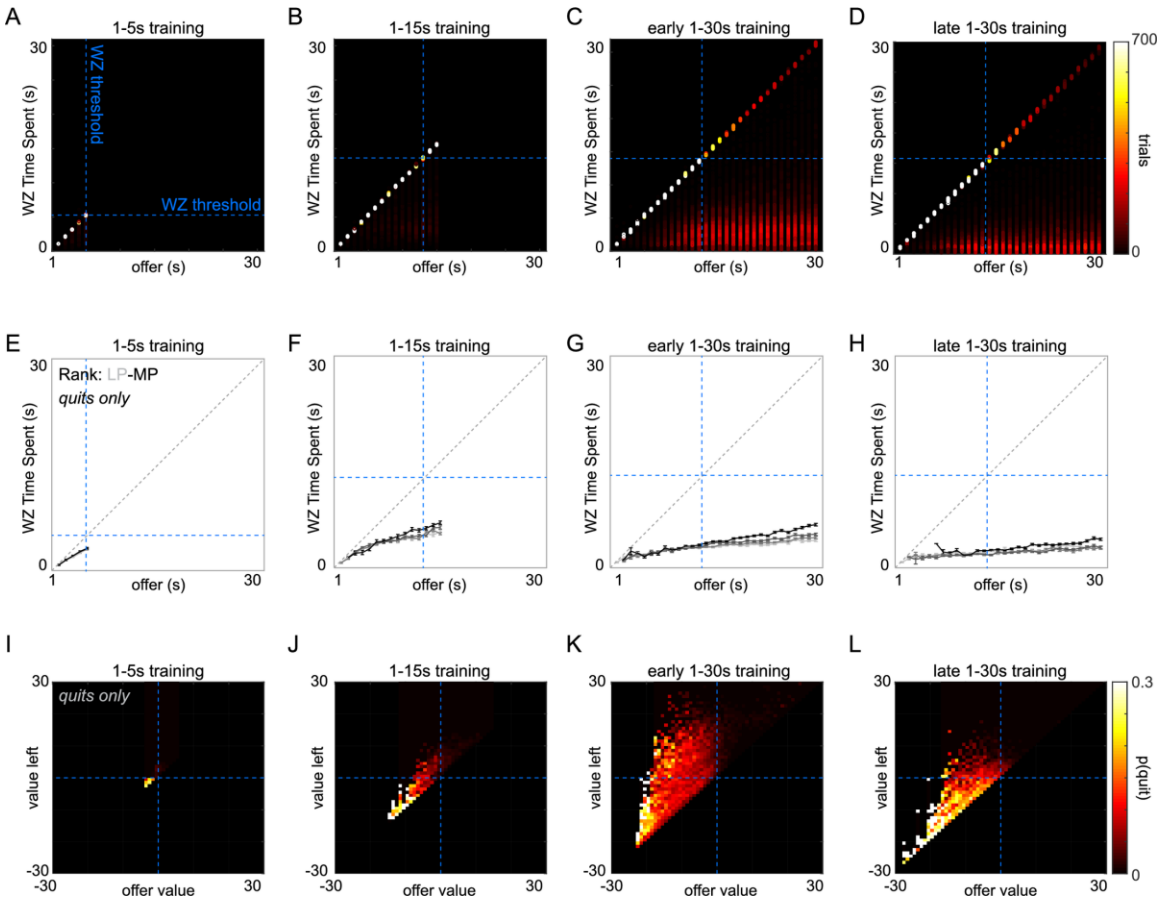
Calculating time spent in the wait zone before quitting across blocks of training reveal mice generally quit relatively quickly regardless of offer cost (Figure 2.10A-D). Following the transition to the 1-30s offer block, I found that mice took significantly longer to decide to quit in the wait zone in more preferred restaurants (Figure 2.3K, Figure 2.10E-H, F=200.94, p<0.0001). These data suggest that mice were more reluctant to change their minds before quitting more preferred offers, demonstrating an aversion to quit reminiscent of the aversion to leave while lingering. As mice learned to forage more economically efficiently in the wait zone during the 2wk adaptation period re-normalizing food intake and reinforcement

Figure 2.9: Distinct decision strategies separately become efficient in the offer zone and wait zone



(A) Offer zone inefficiency ratio. Offer value (VO) = wait zone threshold [WZ] – offer cost [O]. Probability of entering negatively valued offers relative to the probability of skipping negatively valued offers. Horizontal dashed line indicates equivalent 1:1 ratio of entering vs. skipping negatively valued offers. (B) Wait zone inefficiency ratio. Value of time left in countdown at the moment of quitting (VL) = wait zone threshold [WZ] – countdown time left [TL]. Probability of quitting negatively valued offers when VL was positive relative to when VL was still negative. Horizontal dashed line indicates equivalent 1:1 ratio of quitting inefficiently vs. efficiently. Data are presented as the cohort's (N=31) daily means (±1SE) across the entire experiment. Color code on the x-axis reflects the stages of training (offer cost ranges denoted on the top of panel A). Vertical dashed lines (except pink) represent block transitions. * on the x-axis indicates ratio significantly greater than 1:1 immediately following the 1-30 block transition, otherwise * indicates gradual significant changes within the 1-30s block during either the 2wk adaptation period or pink epoch.

Figure 2.10: Economic characterization of wait zone strategy across stages of training



(A-D) Histogram of wait zone events as a function of time spent waiting and as a function of offer cost. Diagonal unity line (time spent waiting = offer cost) represents earned trials while the remaining data points represent quit decisions. Horizontal and vertical dashed lines represent average wait zone thresholds across the 1-5s block (A), 1-15s block (B), early 1-30s block (C, first 5 days), and late 1-30s block (D, last 5 days). (E-H) Average time spent waiting before quitting as a function of offer cost split by flavor ranking. (I-L) Histogram of quit decisions as a function of offer value and value left at the moment of quitting. Offer value (VO) = wait zone threshold – offer cost. Value of time left in countdown at the moment of quitting (VL) = wait zone threshold – countdown time left.

rates, the economic efficiency of quit decisions can be examined by comparing time spent before quitting against time remaining in the countdown at the moment of quitting, relative to wait zone thresholds (Figure 2.10I-L). Wait zone inefficiencies depicted in Figure 2.9B following the transition to 1-30s offers are reflected in Figure 2.10K (upper left quadrant). Interestingly, I found that mice were more inefficient in more preferred restaurants and showed more difficulty learning to forage efficiently in more preferred restaurants (Figure 2.3P, F=3.23, p<0.05). This suggests that the conditioned-place-preference-like effect described previously in post-earn lingering time (where no overt reward was being sought), could be related to the effect during wait zone countdowns showing an aversion to quit a reward-associated context, where the higher value of more preferred restaurants strongly opposed a learning processes to adapt economically advantageous foraging strategies.

Previous reports in rats on other neuroeconomic foraging tasks too demonstrate an aversion to leave potential reward opportunities, which can add weight to reward valuations and actually contribute to sub-optimal decisions (Wikenheiser et al. 2013; Carter and Redish 2016). Consistent with previous reports, I found here that mice too behaved largely sub-optimally in this variant of the Restaurant Row task, particularly due to added weight in aversions to leave that scaled with subjectively preferred flavors. First, it should be worth mentioning that it can be argued that the fact that mice would be willing to work differently for various flavored pellets of the same caloric value at the expense of not maximizing total food intake is overtly sub-optimal in and of itself. This aside, taking into account idiosyncratic differences in subjective flavor preferences, I wanted to characterize how mice went about making economic decisions using their limited time budget normalized to subjective flavor preferences without making assumptions about flavor value. That is, I wanted to quantify how much more sub-optimal did mice behave separate from the fact that an individual mouse's goal might in fact be to earn more of one particular flavor. With each individual's idiosyncratic subjective flavor valuations taken into account, I quantified how mice engaged in a number of economic decision processes that detracted from their total 1hr session time budget in ways that appeared wasteful and effectively reduced maximum potential earnings within a given flavor, even after taking into account their subjective flavor preferences.

I generated a computer-model that simulated Restaurant Row sessions based on each animal's observed behaviors in order to calculate maximal predicted pellets a mouse could earn in each given restaurant on a given day. I used each animal's daily wait zone thresholds for each restaurant to base the model around individual differences in idiosyncratic subjective flavor preferences. However, I instructed the model to minimize time-expenditure inefficiencies and used each animal's best daily behavioral capabilities to predict maximum yield (i.e., bottom quartile offer zone reaction time, no quits, minimal consumption and lingering time, and minimal between-restaurant travel time). Because differences on most of these metrics exist along the least-to-most preferred flavor ranking axis (e.g., aversions to quit or linger that scaled with ranking), I expected to find asymmetries in sub-optimal behavior across the differently ranked restaurants.

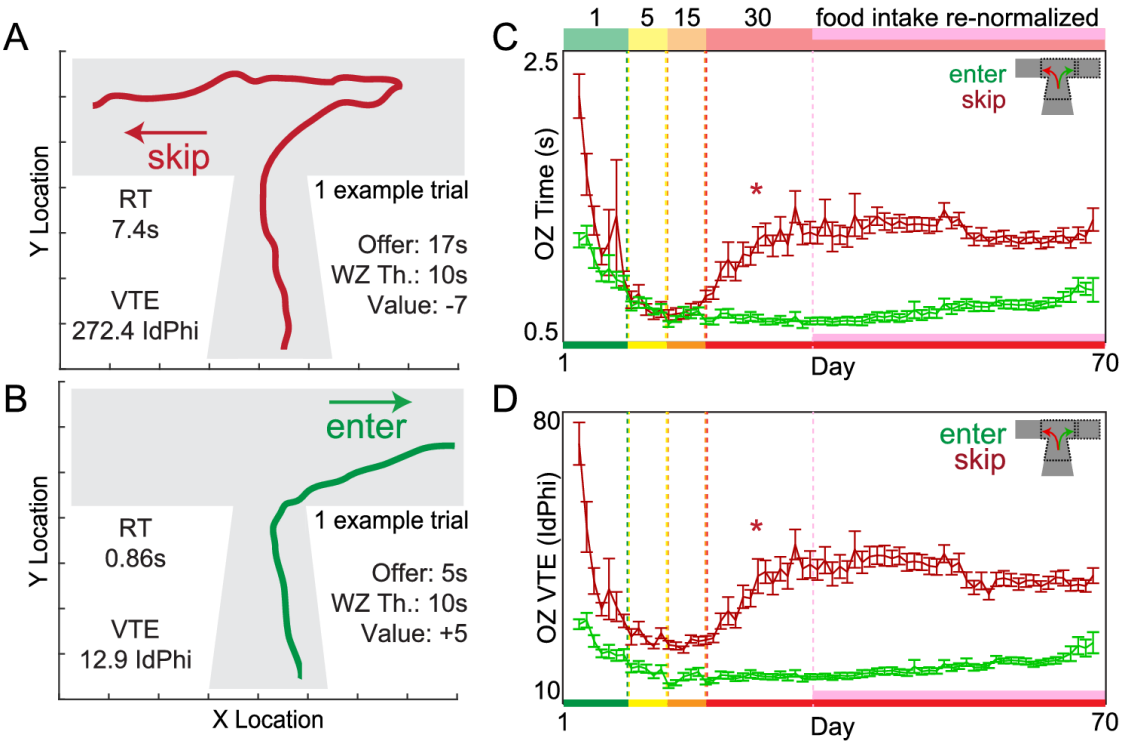
I was surprised to find that mice were more sub-optimal in less preferred restaurants than in more preferred restaurants across the entire experiment (Figure 2.3Q, F=491.22 p<0.0001) even though mice tended to engage in more "wasteful" behaviors in higher preferred restaurants. Perhaps more telling and more expected, I found upon transitioning to the 1-30s offer block, there was a significant interaction between flavor rankings across days on sub-optimal performance (Figure 2.3Q, F=13.57 p<0.0001). Mice became more sub-optimal in the most preferred restaurant immediately following the block transition while becoming gradually more optimal in less preferred restaurants (more robust changes in lesser preferred restaurants) over the subsequent two weeks. The immediate change in optimal performance in the most preferred restaurant (decreased optimality) coincided with the immediate loss in food intake mice experienced while the slower change in optimal performance in lesser preferred restaurants (increased optimality) coincided with the intermediate foraging strategy learning that took place to adapt to a reward-scarce environment and re-normalize overall food intake and reinforcement rates back to levels similar to baseline in reward-rich environments. This is reminiscent of notions of demand elasticity theory and ways in which stubborn pre-cost-change decision policies for luxury goods that remain unchanged in the face of an increasingly reward-scarce environment can lead to immediately drastic changes in optimality.

After mice successfully restored food intake and reinforcement rates by refining a foraging strategy, I found a distinct, delayed phase of additional learning that took place with prolonged training in the absence of any further changes in food intake (pink epoch, Figure 2.2A, F=1.82, p=0.18), reinforcement rates (pink epoch, Figure 2.B, F=0.01, p=0.95), or laps run (pink epoch, Figure 2.C, F=1.54, p=0.21). The proportion of enter-then-quit decisions decreased over the remainder of the experiment (Figure 2.E, F=159.30, p<0.0001) as mice learned to reject offers in the offer zone that they were unwilling to remain committed to once in the wait zone (Figure 2.6D,H). This is reflected in a decrease in offer zone thresholds until they were in register with wait zone thresholds by the end of the experiment (pink epoch, Figure 2.2F, offer zone change: F=812.40, p<0.0001; offer zone vs. wait zone at day 70: F=0.17, p=0.68). As a result, mice learned to skip more often in the offer zone (pink epoch, Figure 2.2D, F=116.85, p<0.0001).

I calculated the economic efficiency of offer zone decisions by measuring the likelihood of skipping offers above wait zone thresholds relative to the likelihood of entering offers above wait zone threshold and found that offer zone decisions became more efficient *only* during the pink epoch (Figure 2.9A, F=474.94, p<0.0001). This onset of offer zone efficiency increase was marked by a clear reversal and onset of decrease in quit frequency and onset of decrease in offer zone thresholds (Figure 2.2E-F). As a result, the proportion of session budget allocated to quit events declined back to baseline levels (pink epoch, Figure 2.4A, budget quitting change: F=1639.61, p<0.0001, day 70 compared to baseline: F=0.17, p=0.68). The only change observed in average time spent per decision across decision processes during this phase of learning was in offer zone time, which increased over extended training as skip frequency increased (pink epoch, Figure 2.4B, offer zone time: F=490.14, p<0.0001; wait zone quit time: F=0.10, p=0.75; earn time: F=0.11, p=0.74; linger time: F=0.73, p=0.39; travel time: F=0.01, p=0.94).

Upon closer examination of offer zone behaviors (Figure 2.11), I found marked changes following the 1-30s transition in skip decisions but not in enter decisions. I calculated the reaction time from offer onset until either a skip or enter decision was made. I also tracked each animal's X-Y-location path trajectory

Figure 2.11: Development of deliberative behaviors during principal offer zone valuations



(A-B) Example x-y-locations of a mouse's path-trajectory in the offer zone (wait zone not depicted) over time during a single trial (from day 70). (A) Skip decision for a high delay offer. The mouse initially oriented toward entering (right) then ultimately re-oriented to skip (left). Wait-zone threshold minus offer captures the relative subjective "value" of the offer. Negative value denotes an economically unfavorable offer. (B) Enter decision for positively valued offer; rapid without re-orientations. This offer zone trajectory pattern is indistinguishable from enter-then-quit decisions for negatively valued offers. (C) Average offer zone reaction time split by enter vs. skip decisions across days of training. (D) Average offer zone vicarious trial and error (VTE) behavior split by enter vs. skip decisions across days of training. Data are presented as the cohort's (N=31) daily means (±1SE) across the entire experiment. Color code on the x-axis in (C-D) reflects the stages of training (offer cost ranges denoted on the top of panel C). Vertical dashed lines (except pink) represent block transitions. * indicates gradual significant changes within the 1-30s block during the 2wk adaptation period.

as they passed through the offer zone. From this, I could capture the degree to which animals interrupted smooth offer zone passes with "pause and look" re-orientation behaviors, known as "vicarious trial and error" (VTE). The physical "hemming and hawing" characteristic of VTE is best measured by calculating changes in velocity vectors of discrete body x-y positions over time as dx and dy. From this, I can calculate the momentary change in angle, Phi, as dPhi. When this metric is integrated over the duration of the pass through the offer zone, VTE is measured in the offer zone as the absolute integrated angular velocity, or IdPhi, until either a skip or enter decision was made (Figure 2.11A-B, day 70 examples path traces).

VTE is a well-studied behavioral phenomenon that reveals on-going deliberation and planning during moments of embodied indecision, supported by numerous electrophysiological experiments reporting concurrent neural representations of possible future outcomes compared serially (Papale et al. 2016; Redish 2016; Muenzinger 1956; Tolman 1939; Johnson and Redish 2007). If VTE is thought to represent episodes of deliberation, planning, and indecision, it would seem that such processes would particularly appear on a neuroeconomic task especially in high conflict scenarios in which mice might be torn between tough, competing options. In this variant of the Restaurant Row task, high-conflict scenarios emerge particularly when mice transition into a reward-scarce environment and offers for highly preferred flavors are made available but at expensive costs. Thus, on this task, I can operationalize the conflict between giving into taking such offers vs. knowing better to forage elsewhere for smarter alternative. A key to interpreting competing valuations in our task during decision-conflict between forward-looking planning and immediate desire-driven responding is the presence or absence of this critical behavioral metric, VTE, which has extensively been studied in a series of "proof of principle" publications (Redish 2016).

In 2007, Johnson and Redish discovered that during VTE, hippocampal representations swept forward along the path of the animal, alternating between potential goals (Johnson and Redish 2007). This key result has been replicated several times. We know that these sequences align to hippocampal theta cycles (Gupta et al. 2012). That is, they are "theta sequences." However, the sequences during VTE sweep farther than during normal navigation (Gupta et al. 2012). The sequences proceed all the way to the goal (Gupta et al. 2012). If

an animal is going to run past one goal to another one, the sequences run farther to the second goal (Papale et al. 2016). They reflect indecision in the animal. An animal that knows where to go does not show VTE and the sequences only sweep forward to the goal the animal is actually going to go to (Papale et al. 2016; Johnson and Redish 2007; Johnson et al. 2007).

Furthermore, neurophysiologically, during VTE, reward-related representations appear in the nucleus accumbens (ventral striatum) and in the orbitofrontal cortex (Steiner and Redish 2012; van der Meer and Redish 2009a; 2009b). Both of these results have been replicated (Stott and Redish 2014). These data suggest that there is an evaluation going along with the prediction in hippocampus. Neurophysiologically, we know that there is a triple dissociation between hippocampus (sweeps during VTE), ventral striatum (reward representations during VTE), and dorsal striatum (no extra activity during VTE, but slowly learned situation-action pairs, van der Meer et al. 2010). As animals develop regular paths and VTE goes away, the dorsal striatum develops "task-bracketing" wherein activity appears at the start of the ballistic journey (Smith and Graybiel 2013). This result has been replicated (Regier et al. 2015). In both of these papers, VTE is negatively correlated to the striatal task-bracketing.

Behaviorally, VTE occurs during times when the animal knows the structure of the world, but doesn't know what to do on it. VTE occurs when the animal is indecisive about goals and when contingencies change (Regier et al. 2015; Schmidt et al. 2013; Amemiya and Redish 2016; Steiner and Redish 2012). Manipulations that force flexibility in tasks lead to an increase in VTE, while manipulations that force regularity in paths lead to a decrease in VTE (Gardner et al. 2013). Finally, on tasks able to differentiate decisions that require planning (sometimes called "model-based") from decisions that reflect cached values (sometimes called "model-free"), VTE occurs when the decisions show planning (model-based) and disappear when the decisions reflect cached values (model-free, Gardner et al. 2013; Schmidt et al. 2013).

In this task, in relatively reward-rich environments, offer zone reaction time became more rapid (green-yellow-orange epochs, Figure 2.11C, $F=157.78$, $p<0.0001$) and path trajectories measured by IdPhi became

more swift and stereotyped (green-yellow-orange epochs, Figure 2.11D, F=150.19, p<0.0001) as mice learned the structure of the task and made ballistic decisions. As these offer zone decisions became ballistic in reward-rich environments as mice were developing subjective flavor preferences, mice also developed rapid default decision responses in each restaurant in doing so. That is, all other things being equal (e.g., all offers were 1s only in the green epoch), the only information differences between restaurants were visual contextual cues that signaled the identity of the flavor of that restaurant. Thus, as animals approached a given restaurant's offer zone after leaving the previous restaurant, mice grew accustomed to release prepotent ballistic decisions to either enter or skip depending on the ranking of that restaurant. For instance, mice grew accustomed to entering and earning nearly 100% of offers in their most preferred restaurants, while skipping the majority of offers in their least preferred restaurants (Figure 2.3A-D, green-yellow-orange epochs – relatively reward-rich environments).

This dichotomy in behavioral responses between least and most preferred restaurants is apparent in a number of economic processes measured on the Restaurant Row task. With regard to developing restaurant-specific default decisions in reward-rich environments, offer zone reaction time grew fastest to enter in most preferred restaurants (Figure 2.3G, F=1076.04 p<0.0001) and fastest to skip in least preferred restaurants (Figure 2.3H, F=70.73 p<0.0001) over the course of the first 17 days of this experiment. Similarly, over the course of the first 17 days of this experiment, VTE was lowest (most stereotyped and least indecisive) when entering in most preferred restaurants (Figure 2.3I, F=592.95 p<0.0001) and when skipping in least preferred restaurants (Figure 2.3J, F=219.93 p<0.0001). Taken together, default ballistic responses to either enter or skip were signaled in the offer zone of restaurants via contextual spatial cues while cost information signaled via tone pitch was irrelevant in reward-rich environments (Figure 2.6A-B,E-F).

However, in a reward-scarce environment, skip reaction time (Figure 2.11C, F=92.00, p<0.0001) and skip VTE (Figure 2.11D, F=117.80, p<0.0001), began to increase following the transition to 1-30s offers. Note the frequency of skip decisions were still relatively rare during the early 1-30s period and did not become more frequent until after food intake and reinforcement rates were restored for the remainder of the

experiment, during which skip time and VTE remained elevated and stable (pink epoch, Figure 2.11C, skip time: F=2.21, p=0.14; Figure 2.11D, skip VTE: F=0.45, p=0.50). Only then did offer zone thresholds decline (Figure 2.2F) and skip frequency increase (Figure 2.2D). Furthermore, following the transition to 1-30s offers, enter decisions remained fast (Figure 2.11C, F=1.73, p=0.19) with low VTE (Figure 2.11D, F=0.97, p=0.32). Taken together, this suggests that enter decisions during this early 1-30s period, when offer zone inefficiencies remained elevated and when quit frequency increases, can be defined as economically disadvantageous snap-judgements made in the offer zone. This also suggests that these quit decisions could serve as economically advantageous re-evaluations of those poor offer zone judgements.

Next, I further characterized the economic nature and utility of VTE behaviors while skipping and potentially preventing poor offer-zone snap judgements to enter "bad" deals one ought to know better not to accept.

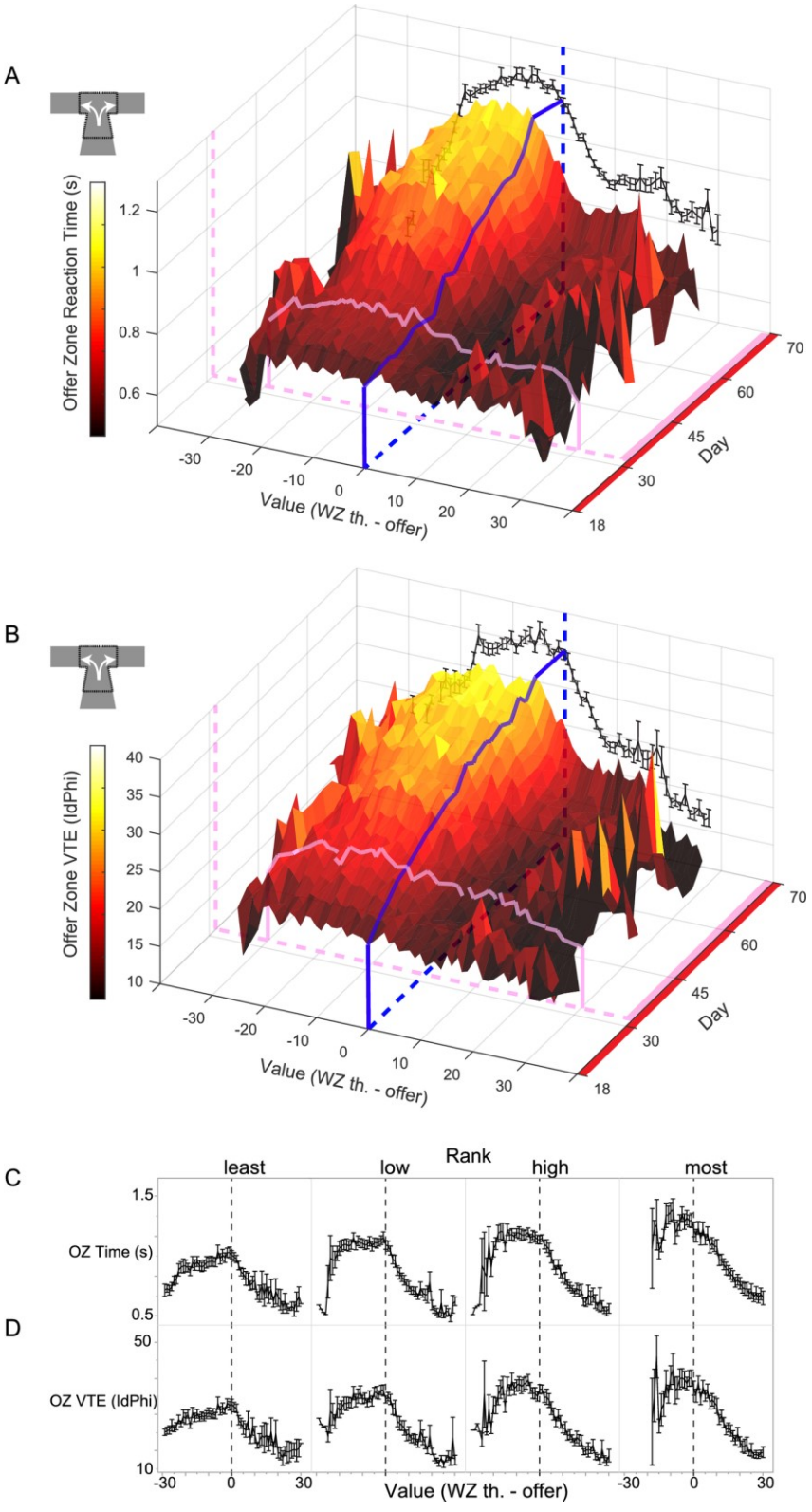
In a reward-scarce environment, reaction time (Figure 2.3H, F=88.78, p<0.0001) and VTE (Figure 2.3J, F=38.74, p<0.0001) was higher when skipping in more preferred restaurants, suggesting decisions to skip expensive offers for desired flavors (and thus oppose stronger initial default decisions to enter by pausing and re-orienting) were more difficult. The fact that skip decisions and not enter decisions displayed increased VTE and more pause-and-reorient behaviors interrupting initial default "almost-enter" decisions suggests that skip decisions recruited an additional delayed process. Such delayed process learned to break initial ballistic snap-judgements and ultimately make the economically advantageous decision to skip expensive offers. This reflects an aversion to skip in the offer zone of preferred restaurants, reminiscent of the aversion to quit in the wait zone and the aversion to leave while lingering at the reward site. Thus, while all three of these conditioned-place-preference-like behaviors appear related, how they manifest and how they change upon transition into a reward-scarce environment a fundamentally different and separable.

I characterized how mice deliberated in the offer zone as a function of offer value. The value of an offer can be operationalized by calculating the different between wait zone thresholds and offer cost, where offers at threshold have a value of zero (indifference point). I found that, over prolonged training, offer zone reaction

time and VTE behavior gradually increased and took on an inverted-U shape that grew stronger with more training, centered near 0 valued offers, but shifted toward negatively valued offers (Figure 2.12A-B). These negatively valued offers that produced peak offer zone decision times and VTE late in 1-30s training reflect the type of expensive offers that, early in 1-30s training, would have been ballistically entered in higher preferred restaurants. This inverted-U shape for offer zone reaction time and VTE that is asymmetrically higher for negatively valued offers compared to positively valued offers was present regardless of flavor preference (Figure 2.12C-D). Taken together, this suggests that high VTE trials for negatively valued offers late in 1-30s training effectively traded enter-than-quit decisions for skips.

In this task, we can take VTE as a sign of indecision and deliberation, and a lack of VTE as a sign of quick, decisive decisions (snap-judgments). In this task, we can reliably detect the difference between VTE and rapid (snap) judgments. Furthermore, I found that when VTE events took place, they did so with delayed onset overriding initial snap judgments in the offer-zone that would have otherwise violated normative economic behavior. This form of delayed deliberative VTE-containing override decisions "rescued" and prevented economic violations from occurring, importantly only when skipping, and could serve as a behavioral operationalization of "knowing better" or "should not" judgments. Sometimes when such slower deliberative VTE process failed to come online, mice accepted expensive offers only to later reverse that initial rapid commitment by quitting in the wait-zone. This indicated that a re-evaluation process can also occur in the wait-zone. Both override-processes in the offer-zone or wait-zone took longer to override in higher preferred restaurants, capturing an increasingly stronger desire-component of these parallel computational processes.

Figure 2.12: Development of deliberative decisions as a function of offer value across training



(A-B) Offer zone reaction time (A) and vicarious trial and error behavior (B, VTE) as a function of offer value (VO = wait zone threshold – offer cost) over days of learning in the 1-30s offer block (red epoch). Blue line represents 0 value trials (where offer = wait zone threshold). Pink line represents onset of food intake and reinforcement rate re-normalization after 2wks of adaptation following the transition to 1-30s offers (pink epoch spans days 32-70). Graphical projection against the back wall displays data presented as the cohort's (N=31) daily means (±1SE) during the last 5 days of training (days 65-70). Z-axis is redundant with color scale for visualization purposes. (C-D) Days 65-70 offer zone time (C) and VTE (D) as a function of offer value split by flavor rank. Vertical dashed black lines represent 0 value trials.

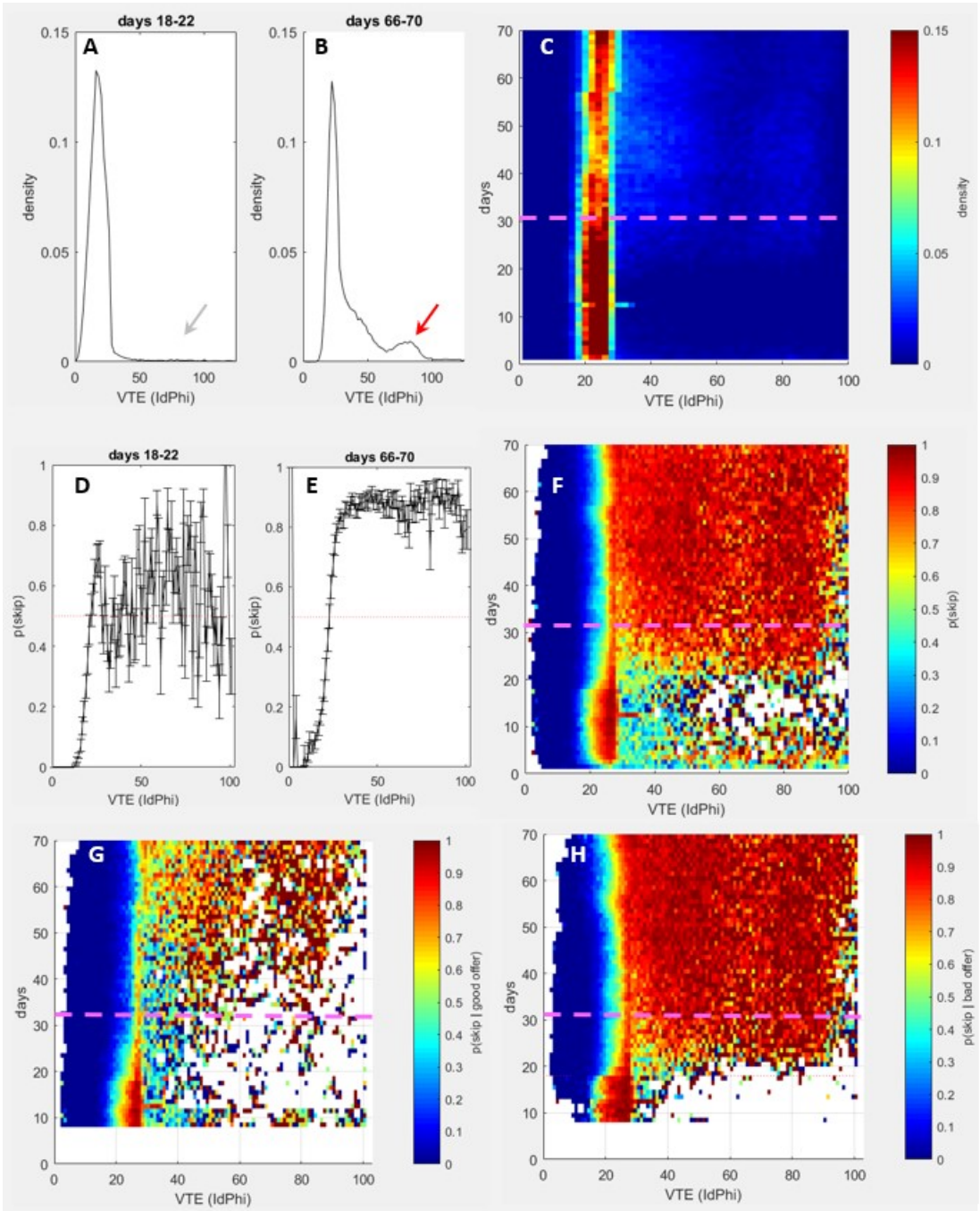
I quantified how the amount of VTE behavior could predict likelihood to make "smart" decisions to skip particularly when offered "bad" deals. Over the course of training, I calculated the distribution of VTE (Figure 2.13A-C). I then calculated the probability of skipping in the offer zone as a function of VTE (Figure 2.13D-F). By looking at offers split by positively vs. negatively valued offers, I found that after extended training, for negatively valued offers, the more mice displayed VTE behavior the more likely they were to skip (Figure 2.13G-H). This suggests that mice enacted deliberative strategies in the offer zone after prolonged training and learned to plan to skip expensive offers that previously would have been rapidly entered then ultimately quit, if mice took the time to deliberate.

When characterizing the economic efficiency of offer zone decisions over time, I found that mice were more inefficient in more preferred restaurants, struggling longest to become efficient in the offer zone of the most preferred restaurant through to the end of the experiment (Figure 2.3O, F=20.72, p<0.0001).

I adopted a signal detection theory approach to further characterize the development of and biases in ability to discriminate offer value over the course of prolonged training (Figure 2.14). I found that receiver operator characteristic (R.O.C.) curves could be used to capture deliberative learning strategies that took place in the offer zone. The area under the R.O.C. curves increased in all restaurants during the 1-30s offer block indicating mice learned to become better value-based signal detectors in the offer zone (Figure 2.14C, F=512.84, p<0.0001). R.O.C. skew captured value-based discriminability error biases in the offer zone (asymmetry towards errors of accepting high cost offers as opposed to rejecting low cost offers) that interacted with subjective flavor preferences (more skew in more preferred restaurants), that did not begin to improve until the pink epoch, and less so in higher preferred restaurants (Figure 2.14D, F=12.94, p<0.0001).

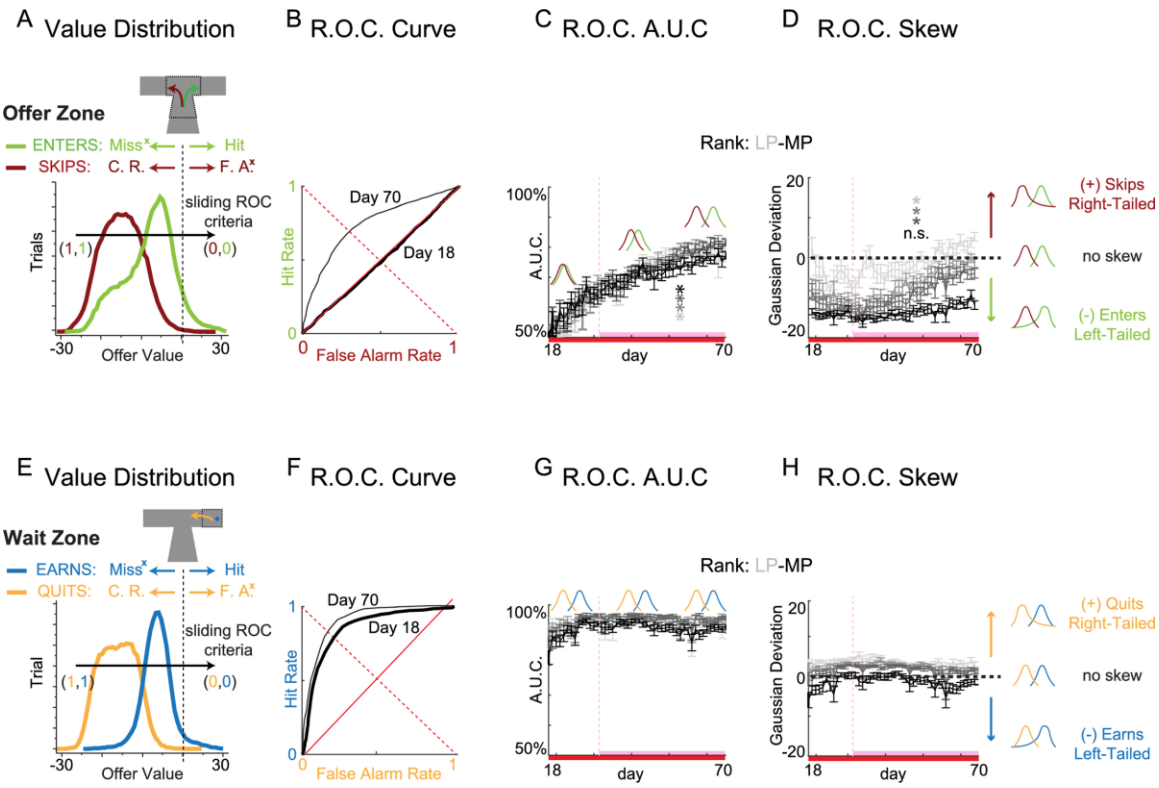
Because skip VTE behaviors begin to increase early in 1-30s training during the time period with food intake and reinforcement is re-normalizing, I wanted to test if any changes in offer-zone behaviors could be contributing to this re-normalization. First, it is worth re-emphasizing that skip behaviors are relatively

Figure 2.13: Characterization of vicarious trial and error (VTE) over training



(A-B) Histogram of distribution of vicarious trial and error (VTE) taken from a span of days early in the 1-30s block (A) and late in the 1-30s block (B, red arrow emphasizes high VTE events). (C) A continuum of the histograms depicted in (A-B) displayed daily across the entire experiment. Horizontal dashed pink line indicates estimated day of food intake re-normalization. (D-E) Probability of skipping in the offer zone as a function of VTE early (D) vs. late (E) in the 1-30s block. (F) A continuum of the p(skip) functions depicted in (D-E) displayed daily across the entire experiment. Note in (D-E), if mice made rapid, highly stereotyped, ballistic choices in the offer zone (low VTE), these decisions were largely enter-decisions. In early training (D), high VTE trials, which were rare, resulted in offer-zone decisions to skip vs. enter that were, at best, made at chance. Skip decisions were also rare early in 1-30s training (Fig. 2D). In late training (E), low VTE trials slightly reduced in frequency while high VTE trials, which were much more frequent, resulted largely skip outcomes, that is, after having heavily deliberated. These decision outcomes did not begin rejecting expensive offers in the offer zone until after food intake re-normalized. (G-H) Split (F) by offers below wait-zone thresholds (G, good deals) and offers above wait-zone thresholds (H, bad deals). (H) drives the majority of (F).

Figure 2.14: Signal Detection Theory characterization of learned value-based discriminability



(A) Offer zone decision distributions as a function of offer value (VO = wait zone threshold – offer cost) split by enter vs. skip decisions. As a function of a sliding receiver operating characteristic (R.O.C.) criterion, R.O.C. curves (B) can be generated by plotting calculated hit rate and false alarm rate pairs at each liberal-to-conservative sliding R.O.C. criterion. Relative to each sliding R.O.C. criterion, hits, misses, false alarms, and correct rejections are characterized by enter vs. skip outcomes for offers whose values lie either to the left or right of the R.O.C. criterion. Economic violations ("X's") represent either "misses" in incorrectly detected criterion-relative negatively valued offers (thus, entering) or "false alarms" (F.A.) in incorrectly detected criterion-relative positively valued offers as criterion-relative negatively valued offers (thus, skipping). Hits represent correctly detected criterion-relative positively valued offers (thus, entering) and correct rejections (C.R.) represent correctly detected criterion-relative negatively valued offers (thus, skipping). Hit Rate = hits / total enters. False Alarm Rate = false alarms / total skips. (B) Offer zone R.O.C. curves changes from being linear (day 18, chance-decision-maker) to bowed-shaped (day 70, good-value-based-signal-detector) quantified by an increase in area under the curve (A.U.C., solid red line indicates chance unity line with 0.5 A.U.C.). (C) Offer zone R.O.C. A.U.C. plotted across days of training in the 1-30s offer block (red epoch) split by flavor ranking. Vertical pink line represents onset of food intake and reinforcement rate re-normalization after 2wks of adaptation following the transition to 1-30s offers (pink epoch spans days 32-70). (D) Offer zone R.O.C. curve skew describes a value bias (tail) of either enter or skip distributions. This is evident in (A) by the (-) left tail of the enter distribution and reflected in (B) by an asymmetry of R.O.C. curve bowedness. Dashed red diagonal line in (B) aids in visualization of R.O.C. curve asymmetry. Gaussian fit peak deviation from this line is quantified in (D). (E-H) reflect same analyses in wait zone decisions. Data (C-D) and (G-H) presented as the cohort's (N=31) daily means (±1SE). * indicate significant change over 1-30s offer block. Not significant (n.s.)

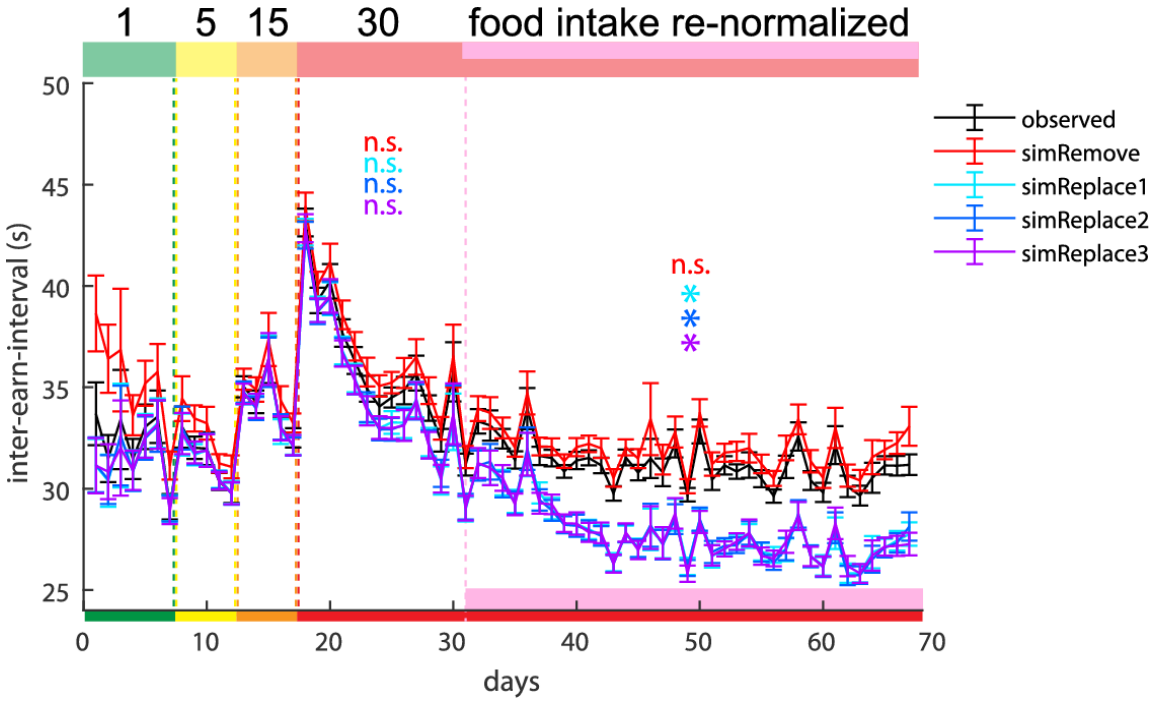
rare events during this epoch of time. Second, choice outcomes in the offer zone during this epoch remain stable and elevated, as mice are apt to accept most offers regardless of cost. During this epoch, decreases in quit time can explain increases in reinforcement rate during re-normalization. Therefore, at first glance, increases in skip time would only oppose an increase in reinforcement rate during re-normalization. Nonetheless, I wanted to test directly how high VTE events could influence reinforcement rates.

I ran computer models to predict how reinforcement rate would be altered if high VTE trials were corrected during Restaurant Row simulations. To do this, I first determined high vs. low VTE by taking a median split of IdPhi distributions across the entire experiment. I generated 4 types of simulations and compared predicted reinforcement rate against the observed reinforcement rate (Figure 2.15). I either entirely removed or replaced high VTE trials with estimated outcomes selected from low VTE trials and found no changes from observed reinforcement rate during the early 1 day -30s epoch (Figure 2.15, F=0.08, p=0.77).

Interestingly, I found effects of high VTE replacement with low VTE estimates compared to observed reinforcements rates and high VTE removal simulations late in 1-30s training. Reinforcement rate was significantly higher (lower inter-earn-interval) in all three replacement simulation variants compared to observed data and predicted data from high VTE removal simulations (Figure 2.15, F=11.98, p<0.0001). Furthermore, reinforcement rates in the high VTE replacement simulations were significantly higher (lower inter-earn-interval) than average reinforcement rates in relatively reward-rich environments (Figure 2.15, first three training blocks, F=33.96, p<0.0001).

Taken together, this suggests that reinforcement rate *could* theoretically improve even better than observed performance only late in training. Thus, when deliberative strategies in the offer zone resulted in changes in offer zone thresholds (i.e., changes in choice outcomes depending on the cost of the offer) and thus offer zone efficiency, it would appear that these decisions were made despite missing out on potentially "better than observed" and even "better-than-before" reinforcement rates. Importantly, this simulation reveals that

Figure 2.15: Controlling for the effects of vicarious trial and error (VTE) on reinforcement rate



Reinforcement rate (inter-earn-interval) is plotted across days comparing observed date (black) vs. four different computer models that simulated what the expected reinforcement rate would be if high VTE trials were adjusted. High vs. low VTE trials were determined by a median split of VTE values taken across the entire experiment. The removal simulation (red) simply removed high VTE trials before reinforcement rates were calculated. The three replacement simulations (cyan, blue, purple) resampled trial outcomes from low VTE trials and differed based on how offer length was resampled on when earned trials were simulated (offer length retained from the high VTE trial, offer length randomly selected from the distribution for low VTE trials, or offer length randomly selected from the uniform range of offers for that block, respectively). These simulations indicate no contributions to reinforcement rate due to high VTE trials during the early 1-30s epoch, despite having an effect late into 1-30s training. Data presented as the cohort's (N=31) daily means (±1SE). * indicate significant difference compared against observed data. Not significant (n.s.).

the excess VTE late in training does, in fact, decrease the reward receipt rate even further below what could, theoretically, be achieved. Again, this would suggest that the decision strategy changes observed in the late 1-30s training are driven by a separate process distinct from that in early 1-30s training, with the latter being VTE-dependent.

By no means can we conclude from these data that sub-optimality on a neuroeconomic task is intrinsically bad nor good. Rather, these data provide economic insight to characterize how adhering to no-longer-sufficient decision policies disrupts reinforcement rates while driving additional learning of new foraging strategies that become more efficient over time while interacting with subjective valuation processes distinct from deliberative strategies that also become more efficient over time. Furthermore, these data make apparent that seemingly "wasteful" behaviors, such as prolonging offer zone deliberation time, which occurs in the latter portion of training (pink epoch), not only affords no changes in overall food intake or reinforcement rate nor changes in optimality (Figure 2.3Q, F=0.05 p=0.82), and perhaps even forgoes potentially better reinforcement rates (Figure 2.15). Perhaps seemingly sub-optimal behaviors that are not directly tied to investing time in the current reward offer (i.e., prolonged deliberation to skip), may be useful in other regards. Taken together, this implies that there must be a "hidden utility" to deliberating or a "hidden cost" to not.

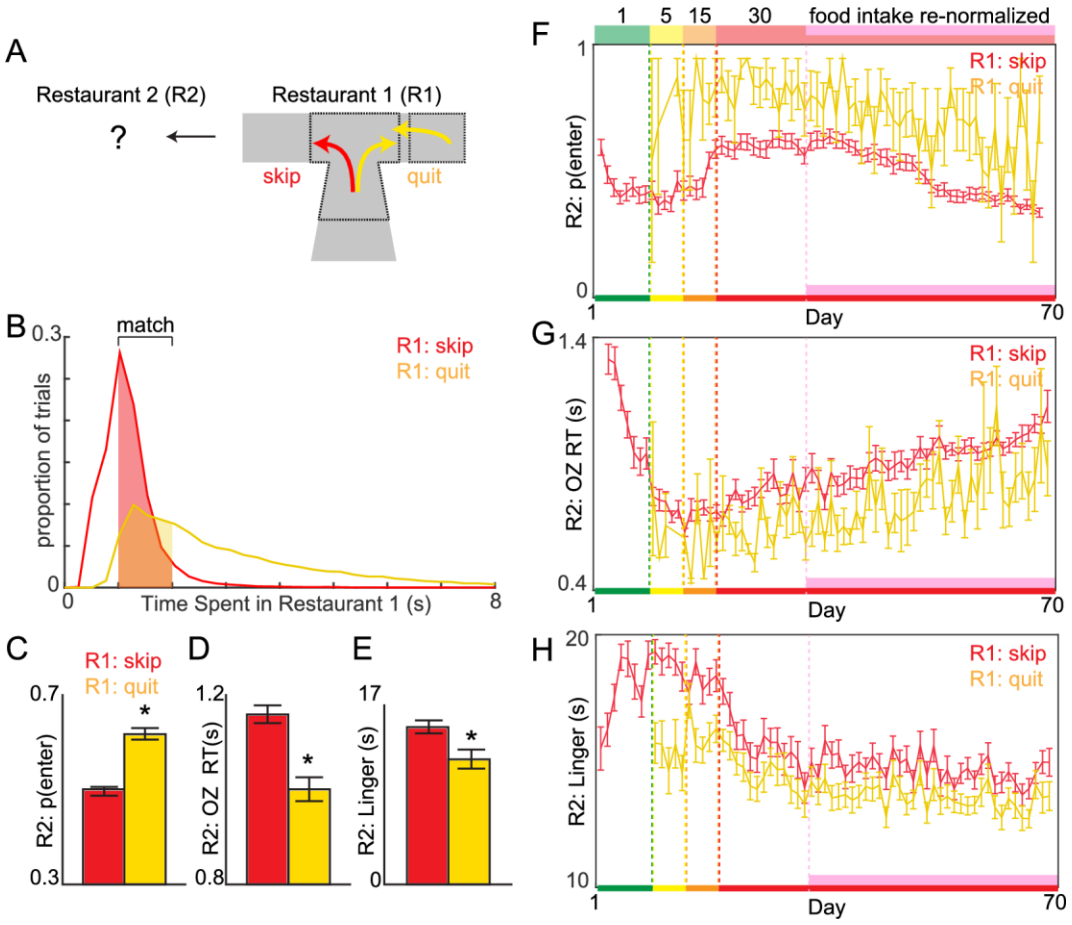
This opens an intriguing question: if the changes that took place with prolonged training did not change the efficiency of food-receipt, and if the only changes after the development of deliberative strategies was a reversal of the increase in quit frequency, what does a reduction in change-of-mind decisions serve these animals? Given that there was no gain in food intake or reinforcement rate nor decrease in energy expenditure, what might the driving force behind this delayed learning process be?

A strength of the Restaurant Row task is its capability of measuring how economic decisions in one trial influence economic decisions in the following trial. This between-trial sequence feature of Restaurant Row captures post-decision-making phenomena, like regret (Steiner and Redish 2014). A key factor in experiencing regret is the realization that a user-driven mistake has been made and that an alternative

response could have led to a more ideal outcome. A change-of-mind quit decision in this novel variant of the Restaurant Row task thus presents an economic scenario where mice take action to opt out of and abandon on-going investments in the wait zone following an economically disadvantageous enter decision. As shown above, quits are economically advantageous re-evaluations of prior snap-judgements made in the offer zone. Thus, quit events reveal a potential economic scenario in which an agent's decision has led to an economically disadvantageous option, whereby a counterfactual opportunity ("should have skipped it in the first place") is realized and could provoke a regret-like experience.

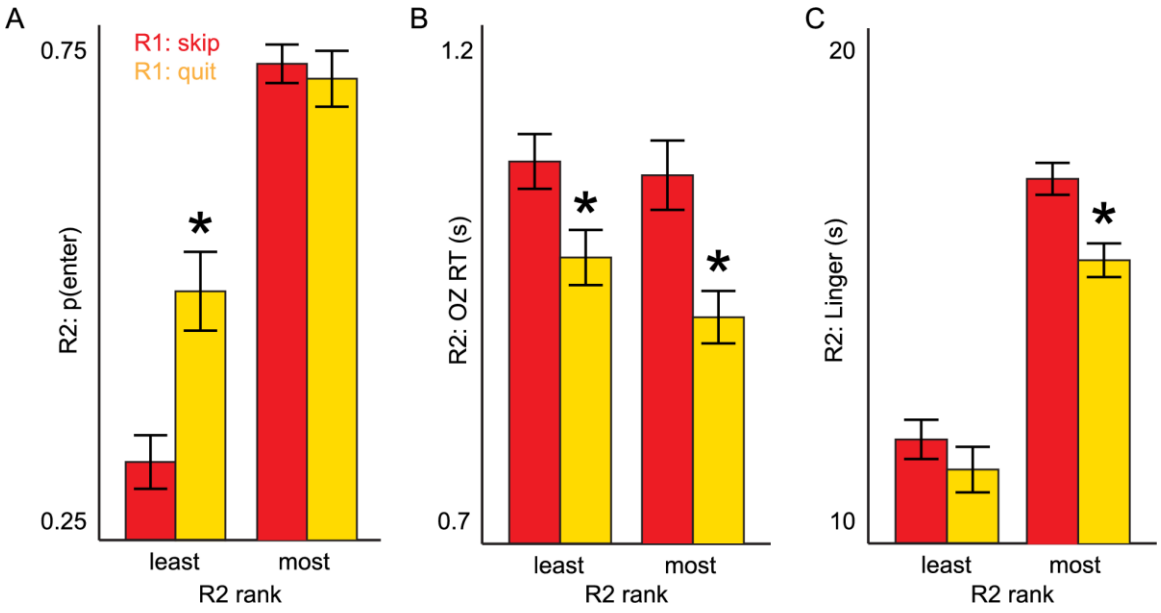
Economic theories of human decision-making have hypothesized that regret adds a negative component to a utility function (Zeelenberg and Pieters 2007; Loomes and Surgden 1982; Patrick et al. 2009; Bell 1982; Coricelli and Rustichini 2010). These theories suggest that an important driving force for human decision-making is the avoidance of future regret (Knutson and Greer 2008; Blanchard and Hayden 2014; Marchiori and Warglien 2008; Frydman and Camerer 2016; Coricelli et al. 2005). In order to test if decisions following enter-then-quit sequences carry added negative utility akin to regret previously demonstrated in Restaurant Row, I examined decision outcomes in the subsequent restaurant encounter following change-of-mind decisions compared to those following skip decisions (Figure 2.16). I compared enter-then-quit events to skip events (Figure 2.16A) that were matched for total time spent in the first restaurant before ultimately turning down the offer and advancing to the subsequent restaurant (Figure 2.16B). For example, I compared a skip decision that used up 2 seconds of offer zone time to an enter-then-quit sequence that used up a total of 2 seconds of combined offer zone and wait zone time. Consistent with previous reports in rats who attempted to make up for lost efforts following regret (Steiner and Redish 2014), I found that mice following quits were more likely to accept offers in the next trial (Figure 2.16C, F=39.26, p<0.0001), did so quickly (Figure 2.16D, F=163.28, p<0.0001), and upon earning subsequent rewards, rapidly consumed food and exited the reward site (Figure 2.16E F=191.89, p<0.0001) compared to trials following skips. Quit-induced effects on subsequent trials existed across the entire experiment (Figure 2.16F-H) and remained even after controlling for flavor preferences (Figure 2.17).

Figure 2.16: Regret-like sequence effects following change-of-mind wait zone re-evaluations



(A) Following either a skip or enter-then-quit decision in Restaurant 1, we characterized behaviors on the subsequent trial in Restaurant 2. (B) Distribution of time spent in Restaurant 1 from offer onset until a skip decision (offer zone time) or quit decision (offer zone time plus wait zone time) was made. To control for the effects of differences in time spent skipping vs. entering-then-quitting in Restaurant 1 on behavior in Restaurant 2, we compared trials matched for resource depletion between conditions. (C-D) Data averaged across the 1-30s offer block. (C) Probability of entering an offer in Restaurant 2 after skipping vs. quitting in Restaurant 1. (D) Offer zone reaction time in Restaurant 2 after skipping vs. quitting in Restaurant 1. (E) Time spent consuming an earned pellet and lingering at the reward site in Restaurant 2 after skipping vs. quitting in Restaurant 1. (F-H) Post-skip vs. post-enter-then-quit sequence data across the entire experiment from (C-E) respectively. Data are presented as the cohort's (N=31) means (±1SE). Color code on the x-axis in (F-H) reflects the stages of training (offer cost ranges denoted on the top of panel F). Vertical dashed lines (except pink) represent block transitions. * indicate significant difference between skip vs. quit conditions.

Figure 2.17: Controlling for flavor preferences in regret-like sequence effects



To control for potential differences in restaurant sequences due to the identity of flavor preferences in Restaurant 2 following quits vs. skips in Restaurant 1, we sorted scenarios such that the Restaurant 2 was always either the least- or most-preferred flavor. (A) Probability of entering an offer in Restaurant 2 after skipping vs. quitting in Restaurant 1. Augmented by quits (increased) vs. skips in only least preferred restaurants. (B) Offer zone reaction time in Restaurant 2 after skipping vs. quitting in Restaurant 1. Augmented by quits (decreased) vs. skips in both least and most preferred restaurants. (C) Time spent consuming an earned pellet and lingering at the reward site in Restaurant 2 after skipping vs. quitting in Restaurant 1. Augmented by quits (decreased) vs. skips in only most preferred restaurants. Data averaged across the 1-30s offer block. * indicate significant difference between skip vs. quit conditions.

I found interactions between subjective flavor preferences in how mice behaved in subsequent restaurants following enter-then-quit decisions compared to skip decisions. That is, following a quit decision, if mice entered their least preferred restaurant next, they demonstrated different immediate post-regret compensatory valuations than if they entered their most preferred restaurant next.

First, mice were more likely to make more rapid decisions in subsequent restaurants following quits regardless if the next restaurant was their least or most preferred flavor (Figure 2.17B, F=37.31, p<0.0001; post-hoc Tukey, least: t=5.39, p<0.0001; most: t=3.28, p<0.01). This suggests mice were more likely to make snap judgements overall following quit-induced regret. Interestingly however, mice were only more likely to enter offers following quits if the next restaurant was their least preferred flavor (Figure 2.17B, F=55.50, p<0.0001; post-hoc Tukey, least: t=9.83, p<0.0001; most: t=0.81, p=0.85). Conversely, mice were only more likely to linger less after consuming earned pellets in the subsequent restaurant if the next restaurant was their most preferred flavor (Figure 2.17B, F=5.10, p<0.05; post-hoc Tukey, least: t=0.89, p=0.81; most: t=3.37, p<0.05). These data suggest that the ways in which mice "made up for lost efforts" and immediately compensated for regret manifested differently based on the subjective value of the flavor of the subsequent restaurant.

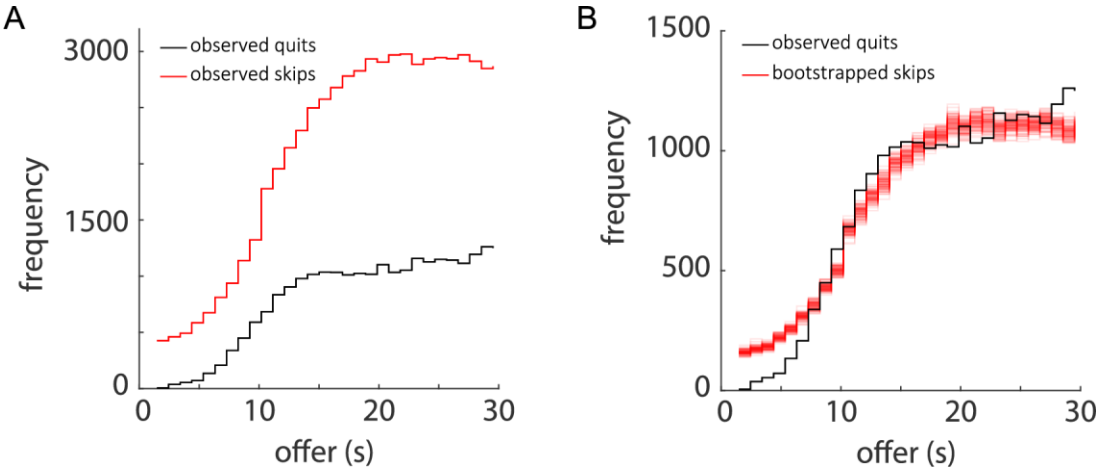
Mice were more willing to accept offers in restaurants they typically defaulted to reject following regret – a finding that may not be observable in most preferred restaurants due to a possible ceiling effect. Mice were also quicker to consume earned pellets and leave the reward site in restaurants they typically lingered at for an extended period of time – a finding that may not be observable in least preferred restaurants due to a possible floor effect. Taken together, these data provide interesting insight when considering how the immediate effects of regret on subsequent decision-making may foster more liberal vs. conservative economic responses, different responses of which may depend on subjective valuation processes of subsequent reward opportunities.

This suggests that enter-then-quit sequences were capable of augmenting subsequent valuations, even when change-of-mind re-evaluations were matched to skip decisions for resource depletion, and even during early stages of training amidst simpler foraging strategies before deliberative strategies developed.

Taken together, on a multiple-week timescale, mice transitioned from a foraging strategy that learned to become efficient (Figure 2.9B) to a distinct deliberative strategy that separately learned to become efficient later (Figure 2.9A). This change in strategy effectively traded enter-then-quit re-evaluative decisions in the wait zone for skip decisions during principal valuations in the offer zone, with no overt benefit other than reducing the frequency of change-of-mind events. Quit events and skip events came from the same distribution of offer lengths (Figure 2.18). I assessed the distribution of offer lengths that comprised the trials in which mice made skip decisions vs. made enter-then-quit decisions to control for potential differences in trial type occurring in restaurant 1 that could confound regret effects in restaurant 2. Because skips occurred more frequently than enter-then-quits, I ran a bootstrapped random re-sampling analysis on the skip events to match the frequency of enter-then-quit trials, and found no differences between these trial types (Figure 2.18).

Based on these data, it seems that not only can a change-of-mind experience have an immediate impact on subsequent valuations, but it can also impact longer-term learning in mice capable of augmenting decision-making strategies. The resulting decision-making strategy appears to be one rooted in deliberation and planning as a means of avoiding future change-of-mind scenarios altogether. In the absence of any additional gains in reinforcement rates, I argue that the utility in developing a deliberative strategy after prolonged training on the neuroeconomic task is rooted in the subjective hidden cost of regret, thus producing a utility benefit to regret avoidance.

Figure 2.18: Visualization of offer length distributions between skip and quit events



(A) Histogram of offer length distributions comparing trials that ended as skips vs. quits from data pooled across animals from days 60-70. (B) Samples were randomly selected from the skip distribution to match the number of samples from the quit distribution. Skip re-sampling was bootstrapped 100 times and replotted in (B). These data indicate both trial types derive from the same offer length distributions.

Discussion

Numerous studies have demonstrated that human individuals develop long-term strategies to avoid future instances of regret (Coricelli et al. 2005; Frydman and Camerer 2016; Camille et al. 2004; Coricelli and Rustichini 2010). This phenomenon is distinct from the ability of regret to drive compensatory augmentations in valuation processes of immediately subsequent opportunities. While the immediate effects of regret have been demonstrated in rodents, long-term regret-avoidance learning however has not been previously observed (Steiner and Redish 2014). Here, I provide support not only for growing evidence that rodents (mice as well as rats) are capable of experiencing regret-like episodes but also that such experiences, separate from and independent of reinforcement maximization, can drive long-term changes in decision-making strategies.

Much of the animal-learning literature has focused primarily on reinforcement maximization as the sole motivator of reward-related learning in decision-making paradigms (Kolling and Akam 2017; Dayan and Niv 2008; Ainslie 1975; Stephens and Krebs 1986). That is, the goal of increasing reward reinforcement rate is thought to underlie animal behavior. Temporal difference error algorithms demonstrate a well-characterized mechanism of reward-maximization-driven motivation in reinforcement learning theory (Stephens and Krebs 1986; Dayan and Niv 2008; Ainslie 1975; Sutton and Barto 1998). Such learning algorithms, supported by neural representations of escalating response vigor and reward expectancies in mesolimbic dopamine systems, update behavioral policies or learn novel contingencies in order to optimize a given cost function and produce maximum reward yield (Holroyd and Coles 2002; Suri and Schultz 1999; Ko and Wanat 2016; Schultz 2017; Schelp et al. 2017; Shizgal 1997). Behavioral and neurophysiological data in both humans and nonhuman animals support a reward maximization theory of learning algorithms.

In the present study, I found evidence of reward maximization learning algorithms as mice progressed from reward-rich to reward-scarce environments and made increasingly efficient wait zone decisions in a self-paced manner on a time-sensitive economic decision-making task during which they earned their only source of food. I also found distinct learning processes separated across space and time in the offer zone that took place on a much longer timescale. I found that mice reduced the frequency of wait zone change-of-mind

decisions by learning to plan ahead in the offer zone, without any additional gain in reinforcement rates or reduction in energy expenditure. Other hypothesized drivers of human learning besides reinforcement maximization and energy expenditure minimization include managing affective states, particularly ameliorating or minimizing negative affect (Kim et al. 2006; Ahn and Picard 2005). Avoiding pain, stress, threat, or anxiety are well-studied motivators in human learning as well as in nonhuman animal fear conditioning or punishment learning paradigms (Krypotos et al. 2015; Kim and Jung 2006). However, in a reward context, negative affect associated with regret and reward-related outcome experiences, while well-characterized in humans, is far less understood in animal learning models of positive reinforcement reward-seeking learning.

The relatively straightforward view of reward-maximization-driven reinforcement learning is challenged by the decision-making phenomena made tractable in these economic decision-making paradigms (Dayan and Niv 2008). Post-decision regret is a well-known example that poses issues for traditional reinforcement learning algorithms dependent on updating stimuli or actions associated with actual experienced reward outcomes (Dayan and Niv 2008). Hypothetical outcomes of forgone alternatives processed during counterfactual thinking that turn out to be better than chosen actions – key in regret – are indeed capable of driving long-term changes in future decision strategies through fictive learning but is a process that has been sparsely studied in nonhuman animals (Abe and Lee 2011; Steiner and Redish 2014; Epstude and Roese 2008; Byrne 2002; Steiner and Redish 2012; Camille et al. 2004; Sommer et al. 2009; Coricelli and Rustichini 2010). Mapping counterfactual outcomes onto corrective actions that could have been taken aids in the development of new decision strategies aimed to avoid regret in the future, yet this is a poorly understood behavioral and neural process.

Change-of-mind behaviors present unique decision-making scenarios, that when assessed on an economic task, can capture the economic advantageous vs. disadvantageous nature of principal valuations and subsequent re-evaluative choices. On this novel variant of the Restaurant Row task, I separate principal valuations (offer zone) from re-evaluative choices (wait zone) across space and time within a single trial.

Two stages of the conflict between wanting highly desired rewards on the one hand versus knowing better to explore other more economically advantageous opportunities on the other hand are captured separately in the offer zone and wait zone. Furthermore, change-of-mind behaviors present a powerful means of studying counterfactual decision processes (Resulaj et al. 2009; van den Berg et al. 2016; Churchland et al. 2008). In the context of the neuroeconomics of regret, a few questions arise: What drives individuals to change their minds? Which decisions might be economically fallible: the original choice, the delayed re-consideration, neither, or both? Why might individuals be reluctant to change their minds, how is this related to regret, and how might this interact with subjective valuation reinforcement learning algorithms?

Change-of-mind decisions occur every day in the real world yet there is the general consensus that many individuals find this unpleasant and are often reluctant to do so, even when its utility is apparent (Kermer et al. 2006; Wilson et al. 2005; Gilbert and Ebert 2002; Roese and Summerville 2005). Imagine the common scenario of a person in a food court during his or her 1hr lunch break deciding which line to wait in – a direct analogue of what I test here in the Restaurant Row task. The decision to back out of waiting in any given line often comes with a sore feeling, even if doing so was an advantageous decision. Conversely, "going down with the ship" describes the sometimes-irrational motivation to refuse overturning a principal judgement and abandoning a partial investment. This is thought to be motivated by a desire in dissonance reduction to avoid being wasteful, admitting mistakes, or challenging one's own beliefs. Thus, following an investment history, it is reasonable to appreciate that progress made toward a goal may be difficult to abandon, doing so may generate a source of cognitive dissonance, and thus the decision to override a principal judgement when re-evaluating continued investment errs on the side of perseveration, however economically irrational that may be. This describes a well-known decision-making phenomenon termed the *sunk cost fallacy*, where the value of continued investment toward reward receipt is inflated as a function of irrecoverable past investments (Arkes and Blumer 1985). As we will see in the coming chapters, mice, rats, and humans on translated variants of the Restaurant Row task all demonstrate the sunk cost effect in the wait zone when making quit decisions as a function of investment history. Thus, quit-induced regret and sunk-cost-driven-perseveration appear to be intimately related here. That is, after making a principal judgement in the offer zone to accept

an offer at a cost higher than subjective value indicates one should (i.e., an initial economic violation of wait zone threshold), subjects are faced with a change-of-mind dilemma torn between irrationally waiting out the expensive offer vs. rationally back-tracking and changing their plans, where affective contributions appear to weigh these options against one another.

In our food court example, the economically-rational decision would be to select a line immediately and to make one's decision while waiting in line. However, this is not what is typically observed – instead, it is far more common for people to deliberate before choosing and investing in any one option, despite the fact that this wastes time planning. Despite re-evaluating an ongoing investment being the economically efficient and rational strategy, this hinges on a high frequency of change-of-mind decisions. After prolonged training in the Restaurant Row task, mice show a shift from the select-and-re-evaluate foraging strategy to the deliberate-first strategy even though it produces no change in reinforcement rate or energy expenditure. Thus, I conclude that mice are capable of learning from regret-related experiences induced by change-of-mind decisions and that they develop a forward-looking deliberative strategy that, although expensive in time and in computational resources, is economically advantageous because regret itself induces a negative utility. Rather than learning to deal with regret, sometimes, mice take the time to plan ahead, and learn to just avoid regret altogether.

– Interlogue –

Chapter 3

The development of distinct valuation algorithms

By taking a neuroeconomic approach, I was able to reveal distinct aspects of decision making in mice through behavior. In the previous chapter, I demonstrated dissociable learning processes that took place in separable decision-making modalities. These separate decision-making modalities – deliberative valuation algorithms and foraging valuation algorithms – were identifiable through discrete behaviors separated across space and time within the same trial. I operationalized these as initial or principal "choose-between" decisions in the offer zone separate from secondary or re-evaluative change-of-mind "opt-out" decisions in the wait zone. This novel variant of the Restaurant Row task is capable of separating these sorts of subjective valuation processes from more general behavioral processes that depend on other forms of learning and memory (e.g., locomotor, auditory, and spatial learning). How these distinct decision-making modalities learned to separately develop over weeks revealed hidden costs and hidden utility associated with each decision system.

In this novel variant of the Restaurant Row task, separate deliberative and foraging valuations took place in the offer zone and wait zone. Importantly, on this task, decisions are interdependent across trials and across days. This critical contingency revealed interesting consequences when mice moved from reward-rich to reward-scarce environments, with an uncompensated budget. Based on neuroeconomic principals and theories of demand elasticity, I discovered several asymmetries in the types of decisions that were made as a function of subjective value and revealed flavor preferences.

Unique cases of economic conflict were revealed, where decision strategies that were previously sufficient in reward-rich environments (i.e., take any offer inconsequentially for desired flavors) were no longer sufficient in reward-scarce environments. This generated interesting high-conflict economic scenarios where tough decisions had to be made for desired although expensive reward opportunities that opposed originally learned valuations (i.e., learn to no longer enter on every trial). Thus, mice learned to override originally learned valuations that "wanted" every reward offer with newer valuations to "know better not to." Aspects of impulsivity and self-control could then be monitored separately in the offer zone and wait zone during high-conflict scenarios.

I found that change-of-mind foraging decisions to quit erroneously accepted expensive offers once in the wait zone become efficient as mice learned to realize that better opportunities lied elsewhere and abandoning the current investment was a worthwhile decision. I also found that this decision came with a hidden cost, not rooted in losses in reinforcement rate, but rather in regret-like negativity that such offers "should have been skipped in the first place." This hidden cost was capable of augmenting immediately subsequent valuations. Perhaps more interestingly, mice learned to develop distinct deliberative strategies in the offer zone that learned to plan ahead during high-conflict economic scenarios. By deliberating, mice could overcome immediately present "wanting" valuations by processing smarter valuations that knew better to skip instead of enter. Therefore, an accept-everything-then-quit foraging strategy, while capable of maintaining re-normalized reinforcement rates, was "revealed more costly than" a plan-first-and-enter-only-if-committed strategy, even after showing that mice theoretically could earn more rewards in the former strategy. Taken together, mice adopted a choose-between deliberative strategy that was capable of bypassing potentially regret-inducing scenarios in the wait zone, thereby conveying this strategy's hidden utility. Thus, deliberation, while time-consuming and computationally intensive, can be worthwhile.

This opens an intriguing question: if change-of-mind decisions, even if economically favorable and the right thing to do, are capable of eliciting feelings of regret and mice are willing to sacrifice increased reinforcement rates just to avoid the negative consequences of experiencing regret by learning to plan ahead, is this the only

way mice can avoid regret? Might mice be vulnerable to avoid feeling regret by inflating the value of not quitting while still in the wait zone?

This calls to a well-known human phenomenon known as the "sunk cost fallacy." The sunk cost fallacy describes the economic bias in which individuals escalate commitment of continued reward pursuit as a function of prior irrecoverable investments already made. As a result, individuals become prone to avoid progress abandonment even if it is the economically advantageous thing to do.

In the next chapter, I will explore the concept of the sunk cost fallacy in this task. Importantly, because this is a well-studied human phenomenon widely thought to be unique to humans, and because there are conflicting reports in non-human animal studies, I take a translational approach across species using the same tasks. Importantly, I continue to carry this concept of multiple, parallel decision-making systems forward and discover novel determinants of the sunk cost fallacy rooted in these separable decision-making systems conserved between species across evolution.

– Original Research –

Chapter 4

Mice, rats, and humans make decisions dependent on perceived "sunk costs," but not while deliberating

Abstract

Sunk costs are irrecoverable investments that should not influence decisions because decisions should be made based on expected future consequences. Both human and non-human animals can show sensitivity to sunk costs, but reports across species are inconsistent. In a temporal context, a sensitivity to sunk costs arises when an individual resists ending an activity, even if it seems unproductive, because of the time already invested. In two novel, parallel foraging tasks, I find that mice, rats, and humans show similar sensitivities to sunk costs in their decision-making. Surprisingly, sensitivity to time invested accrued only after an initial decision had been made. These findings suggest sensitivity to temporal sunk costs lies in a vulnerability distinct from deliberation processes, and that this distinction is present across species.

Chapter reprinted with permissions from *AAAS*, modified from:

Sweis BM, Abram SV, Schmidt BJ, Seeland KD, MacDonald AW, Thomas MJ, Redish AD. 2018c. Sensitivity to "sunk costs" in mice, rats, and humans. *Science* 361: 178-181.

Introduction

Traditional economic theory suggests that decisions should be based on valuations of future expectations that ignore spent resources that cannot be recovered (sunk costs, Thaler 1999). However, extensive evidence finds that humans factor such sunk costs into prospective decisions even when faced with better alternatives (Arkes and Ayton 1999; Höffler 2005). Although early reports claimed that humans are uniquely sensitive to sunk costs, it is becoming increasingly clear that non-human animals exhibit parallel behaviors (Höffler 2005; Arkes and Ayton 1999; Magalhães and White 2016).

Previous non-human animal studies that attempted to model the sunk cost phenomenon yielded conflicting evidence (Arantes and Grace 2008; Magalhães and White 2016). Observational and experimental field studies in swallows, sparrows, mice, and bluegills have found evidence both for and against the sunk cost effect in behaviors relating to parental investment and willingness to care for young (Coleman et al. 1985; Dawkins and Carlisle 1976; Maestripieri and Alleva 1991; Weatherhead 1979; Winkler 1991). Yet in such studies, it has been difficult to disentangle influences of investment history from those of future prospects. Laboratory operant conditioning paradigms in pigeons and rats that control for future expectations when looking at reinforcement learning behaviors have demonstrated that non-human animals showed increased work ethic or sub-optimal perseverative reward-seeking behaviors that indeed escalate with prior investment amount (Magalhães et al. 2012; Pattison et al. 2012). However, these observations often relied on information uncertainty where subjects over-worked in the absence of progress-indicating cues. These observations also often relied on automation or habit-like behaviors (e.g., repetitive lever pressing) driving continued reward-pursuit. Such confounds in non-human animal studies obscure translation to human sunk cost effects, which do not depend on these mechanisms.

Laboratory foraging tasks provide an alternative approach to study decision making using naturalistic behaviors that carry both ecological validity and evolutionary significance and are translatable across species (Kalenscher and van Wingerden 2011). Foraging tasks rely on optimizing reward-seeking under limited resources, making them economic tasks.

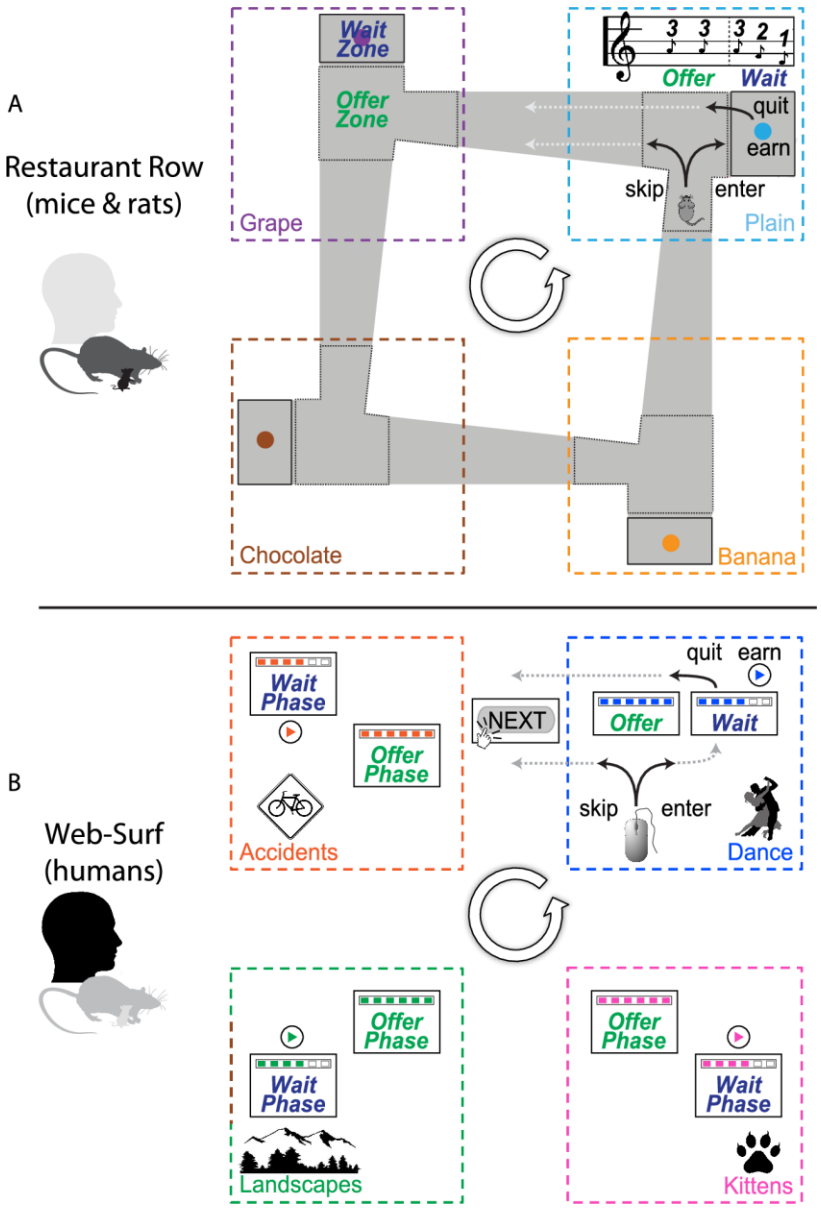
We designed a foraging task in which subjects spend time from a limited time-budget waiting for rewards (Figure 4.1, Restaurant Row, Steiner and Redish 2014; the Web-Surf Task, Abram et al. 2016).

Methods

This project was a joint collaboration across three separate labs (Redish, MacDonald, and Thomas labs), as such, we analyzed data that had been collected under various testing conditions (see Table 4.1 for chronology). The Restaurant Row Task was first developed in the Redish lab for use in rats using the original task variant without an offer zone (Cohort 1). In collaboration with the MacDonald lab, this task was translated for humans in the variant of the Web-Surf Task without an offer phase (Cohort 2) that mirrored the initial rat variant. Shortly after, collaboration with the Thomas lab translated this task for mice. At this time, the separate offer zone was added to deconstruct stages of decision-making within trial (Cohort 3). The initial goal was to glean more discrete metrics from an already behaviorally rich task. Mice generally took longer to train than rats (and humans), but after over a year of in-depth behavioral analyses modeling separate economic valuation algorithms in different aspects of behaviors in the offer zone and wait zone, an extended discussion on the sunk-cost fallacy grew as initial discoveries were made and quantitative control analyses were being designed. Moving forward, this inspired running new cohorts of rats and humans on this novel task variant with separate offer zones and wait zones in the Redish and MacDonald labs (Cohorts 4 & 5, respectively). The goal of Cohorts 4 & 5 was to match the mouse protocol (Cohort 3) as closely as possible. The table below compares and contrasts the cohorts tested across species and task variants. This table also includes the chronological order of cohorts.

There are several differences across Cohorts 1-6 including different experimenters that ran the test subjects or different maze apparatuses used between rodents (Cohort 1 rats were tested in an open apparatus without

Figure 4.1: Cross-species task schematics



(A) Restaurant-Row Task: Food-restricted rodents were trained on a maze encountering serial offers for flavored rewards in four "restaurants." Each restaurant contained a separate offer zone and wait zone. Tones sounded in the offer zone; fixed tone pitch indicated delay (1-30s, randomly selected) rodents would have to wait in the wait zone. Tone pitch descended in the wait zone during delay "countdown." Rodents could quit the wait zone for the next restaurant during the countdown, terminating the trial. (B) Web-Surf Task: Humans performed an analogous 30min computer-based foraging paradigm encountering serial offers for short entertaining videos from four "galleries." A static "download bar" appeared in the offer phase indicating delay length (1-30s, randomly selected) that did not begin downloading until after entering the wait phase. Downloads could be quit during the wait phase.

Table 4-1: Cohort information in chronological order of data collection

	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Cohort 5	Cohort 6
Date	Spring 2014	Spring 2014	Fall 2016	Spring 2017	Spring 2017	Fall 2017
Species	Rat	Human	Mouse	Rat	Human	Mouse
Breed	Brown-Norway	undergraduates	C57BL6J	Fisher Brown-Norway	undergraduates	C57BL6J
Sample Size & Sex	22 (M) & 0 (F)	4 (M) & 13 (F)	32 (M) & 0 (F)	4 (M) & 6 (F)	24 (M) & 41 (F)	32 (M) & 0 (F)
Age	8-12 months	19.63 years (mean)	13 weeks	6-10 months	20.23 years (mean)	13 weeks
Task Variant	wait zone only	wait phase only	offer zone + wait zone	offer zone + wait zone	offer phase + wait phase	offer zone + wait zone
Experimenters & Gender	1 (M) & 1 (F)	3 (M) & 5 (F)	2 (M) & 3 (F)	2 (M) & 2 (F)	0 (M) & 6 (F)	3 (M) & 3 (F)
Length of Training	20+ days	5 minutes	70+ days	20+ days	5 minutes	70+ days
Food Deprivation	>80% free weight	N / A	>80% free weight	>85% free weight	N / A	>90% free weight

The sequence of experiments that took place across labs inspired and informed subsequent experiments. The similarities and differences across species in fact reflect a strength that our work capitalizes on, demonstrating robustness of main effects despite these variations as well as harnessing interesting differences in certain effects because of between-cohort differences. The approach and design we used makes this naturalistic foraging task easily expandable to a wide variety of populations, including different ages or patient populations in humans (*undergrads are an interesting breed in their own right), as well as other species, beyond rodents.

walls. Cohort 3 mice were tested in an enclosed apparatus with walls after initial failed attempts building an open apparatus without walls during the translation processes. Therefore, Cohort 4 rats were tested in an enclosed apparatus with walls to match mice.) We would argue that in spite of these differences, we still observed robust effects. The fact that the effects are so remarkably similar across species and task variations despite the experimental differences increases our confidence in our results.

Mice

32 C57BL/J6 male mice, 13 weeks of age, were trained in Restaurant Row. Mice were single-housed in a temperature- and humidity-controlled environment with a 12-hr-light/12-hr-dark cycle with water ad libitum. Mice were food restricted to a maximum of 85% free feeding body weight and trained to earn their entire day's food ration during their 1-hr Restaurant Row testing session. A replication cohort of an additional 32 mice were run. These mice were intentionally food restricted to a lesser extreme (a maximum of 90% free feeing body weight). All experiments were approved by the University of Minnesota Institutional Animal Care and Use Committee.

Rodent flavored-pellet training (mice)

Mice underwent 1 week of pellet training prior to the start of being introduced to the Restaurant Row maze. During this period, mice were taken off of regular rodent chow and introduced to a single daily serving of BioServ full nutrition 20mg dustless precision pellets in excess (5g). This serving consisted of a mixture of chocolate-, banana-, grape-, and plain-flavored pellets. Next, mice (hungry, before being fed their daily ration) were introduced to the Restaurant Row maze 1 day prior to the start of training and were allowed to roam freely for 15min to explore, get comfortable with the maze, and familiarize themselves with the feeding sites. Restaurants were marked with unique spatial cues. Feeding bowls in each restaurant were filled with excess food on this introduction day.

Restaurant Row procedure (mice)

Task training was broken into 4 stages. Each daily session lasted for 1hr. At test start, one restaurant was randomly selected to be the starting restaurant where an offer was made if mice entered that restaurant's T-shaped offer zone from the appropriate direction in a counter-clockwise manner. During the first stage (day 1-7), mice were trained for 1 week being given only 1s offers. Brief low pitch tones (4000Hz, 500ms) sounded upon entry into the offer zone and repeated every second until mice skipped or until mice entered the wait zone after which a pellet was dispensed. To discourage mice from leaving earned pellets uneaten, motorized feeding bowls cleared any uneaten pellets upon restaurant exit. Left over pellets were counted after each session and mice quickly learned to not leave the reward site without consuming earned pellets. The next restaurant in the counter-clockwise sequence was always and only the next available restaurant where an offer could be made such that mice learned to run laps encountering offers across all four restaurants in a fixed order serially in a single lap. During the second stage (day 8-12), mice were given offers that ranged from 1s to 5s (4000Hz to 5548Hz, in 387Hz steps) for 5 days. Offers were pseudo-randomly selected such that all 5 offer lengths were encountered in 5 consecutive trials before being re-shuffled, selected independently between restaurants. Again, offer tones repeated every second in the offer zone indefinitely until either a skip or enter decision was made. In this stage and subsequent stages, in the wait zone, 500ms tones descended in pitch every second by 387Hz steps counting down to pellet delivery. If the wait zone was exited at any point during the countdown, the tone ceased and the trial ended, forcing mice to proceed to the next restaurant. Stage 3 (day 13-17) consisted of offers from 1s to 15s (4000Hz to 9418Hz) for another 5 days. Stage 4 (day 18-70) offers ranged from 1s to 30s (4000Hz to 15223Hz) and lasted until mice showed stable economic behaviors. Early 1-30s and well-trained 1-30s timepoints used in analyses include the first 5 and last 5 days of stage 4. We used 4 Audiotek tweeters positioned next to each restaurant powered by Lepy amplifiers to play local tones at 70dB in each restaurant. We recorded speaker quality to verify frequency playback fidelity. We used Med Associates 20mg feeder pellet dispensers and 3D-printed feeding bowl receptacles fashioned with mini-servos to control automated clearance of uneaten pellets. Animal tracking, task programming, and maze operation was powered by AnyMaze (Stoelting). Mice were tested at the same time every day in a dim-lit room, were weighed before and after every testing session, and were fed

a small post-session ration in a separate waiting chamber on rare occasions as needed to prevent extremely low weights according to IACUC standards (not <85% free-feeding weights).

Rats

10 Fisher Brown-Norway rats (4 male, 6 female), aged between 8-12 months, were trained to run the Restaurant Row task variant with an offer zone. 22 Brown-Norway rats (male), aged between 8-12 months, were trained to run the Restaurant Row task variant without an offer zone. Rats were single-housed in a temperature and humidity-controlled environment and kept on a 12hr light/dark cycle with water ad libitum. Rats were food restricted to a maximum of 80% free feeding body weight and earned their food each day during their 1-hr Restaurant Row session. All experiments were approved by the University of Minnesota Institutional Animal Care and Use Committee.

Rodent flavored-pellet training (rats)

All rats were given 8 days of handling and pellet training prior to being introduced to the Restaurant Row task. Handling consisted of roaming freely on the experimenter's lap for 15-20 minutes daily. For pellet training, rats were taken off ad libitum access to Teklad rodent chow and given 1hr of access to 15g of TestDiet full nutrition 45mg purified rodent tablets in four unique flavors (chocolate, banana, cherry, plain). Rats were allowed to eat freely, either while being handled, in a bedding-free cage, or a combination of both.

Restaurant Row procedure (rats)

Training in the task variant with an offer zone consisted of four training phases. Each session lasted 60min. Rats ran each phase for five days. Phase one consisted of 1s delays at each restaurant. Phase two consisted of randomly selected delays from 1-5s. Phase three included 1-15s offers, and phase four had the final delay range of 1-30s offers. Training in the task variant without an offer zone in a separate cohort of rats was slightly different. These rats were initially trained twice a day in 30min sessions. Training began with 5 days of 1s offers at all feeder sites. Then, the randomized list of delays presented to animals was expanded to 1-2s, 1-3s, 1-4s, and 1-5s delays over 4 consecutive days. Rats then received 10 days of 1-30s delays. Next,

rats switched to once a day 60min testing sessions using 1-30s delays. All delays were randomly selected and varied between day and restaurant. In the task variant with an offer zone, maze contingences were similar to mice described above. In the task variant without an offer zone, delay countdowns began immediately upon entry into the restaurant. Rats ran in a counter clockwise direction, where offers were only triggered if the rats passed through each restaurant in serial order, such that trials would not be triggered when running backwards. After triggering a trial, the next available restaurant where an offer could be made was always the restaurant immediately after the last restaurant triggered, regardless of if an offer was accepted or declined. Rats were run at the same time each day in a very dimly lit room. At the start of the task, rats were always placed on the maze in the same place. Rats were weighed before running the task. If a rat was at or near their 80% weight and did not receive enough food on the track during their running session, they were fed no sooner than a half an hour after completing their running session. Maze operation was done by Matlab. We used 45mg Med Associated feeders to deliver the pellets into in-house 3D printed feeding bowls.

Humans

65 undergraduate students from the University of Minnesota completed the Web-Surf Task (24 male, 41 female, mean age = 20.23 years), and an additional 17 completed the task variant without an offer zone (4 male, 13 female, mean age = 19.63 years); of note, 14 of these 17 subjects represent a subset of the data presented in Abram et al. 2016. Participants received compensation in the form of extra credit towards psychology courses. Ethnicity of subjects included 73% White, 16.5% Asian, 4.5% Black/African American, 2.5% Hispanic/Latino, 0.5 American Indian/Alaska Native, 0.5% Native Hawaiian/Pacific Islander, and 2.5% other. The University of Minnesota Institutional Review Board approved the human study procedures, and all undergraduate subjects provided written informed consent.

Web-Surf procedure

In the task variant with an offer phase, subjects had 30 minutes to travel between four galleries that included video rewards; we used the same categories described in Abram et al. 2016: kittens, dance landscapes, and bike accidents (Abram et al. 2016). Offers were presented in text and with a webpage-like progress bar. When

subjects arrived at a gallery, they first had the option to stay or skip. If they chose to skip, they traveled on to the next gallery and encountered a new offer. If they stayed, they entered the wait phase, and the progress bar begins to count down. At any point before the delay finished, the subject could elect to quit, which again led the process of traveling to the next gallery. If the subject stayed through the entire delay, they were shown a video reward for 4 seconds, after which they rated the video from 1-4 (4 = most enjoyed) according to how much they liked that video. When traveling between galleries, subjects pressed a series "next" buttons as they randomly appeared around a gray screen; numbers were presented in a slightly darker shade of gray to increase task difficulty. For training, subjects completed 3 forced-choice trials to illustrate what happens in the enter, skip, and quit conditions. The forced-choice trials are followed by 8 additional practice trials where subjects make decisions.

In the task variant without an offer phase, subjects similarly had 30 minutes to travel between the same four categories, with offers presented in the same manner. In this variant, the delay began to countdown immediately upon gallery arrival. Subjects then had the option to quit and move on to the next gallery or continue to wait for the delay to finish. Videos were again shown for 4 seconds, and subjects rated each viewed video on a 1-4 scale (4 = most enjoyed). When passing between galleries, subjects clicked a series of "next" buttons that randomly appeared around the gray screen; the buttons were shown in dark gray to blend into the background to increase task difficulty. For training, subjects completed 2 forced-choice trials to illustrate quit decisions, followed again by 8 practice trials where they could decide whether to quit or earn. Regardless of the task variant, subjects ranked the categories by preference (again from 1-4, 4 = most preferred) after the testing session in a post-testing debriefing survey.

Sunk cost data analysis

All data were processed in Matlab and statistical analyses were carried out using JMP Pro 13 Statistical Discovery software package from SAS. The main analysis carried out in the manuscript critical to our findings included computing linear regression models of earning probability in the wait zone as a function of either time remaining in the wait zone or time spent in the offer zone. These analyses compare the slope

coefficients of these regressions interacting across various sunk cost conditions using an ANOVA with p(earn) as the dependent variable and time-remaining x sunk cost condition as factors. Importantly, control analyses were re-calculated to control for subtle skews in different dataset availability distributions between the various sunk cost scenarios. Comparisons of interest included testing regression coefficients of the zero sunk cost condition against each sunk cost condition – data originating from black dataset against color datasets after taking into account proper adjusted control datasets as well as testing regression coefficients of each sunk cost condition against other sunk cost conditions – color data against color data, again, after taking into account proper adjusted control datasets.

Modeling sub-optimality

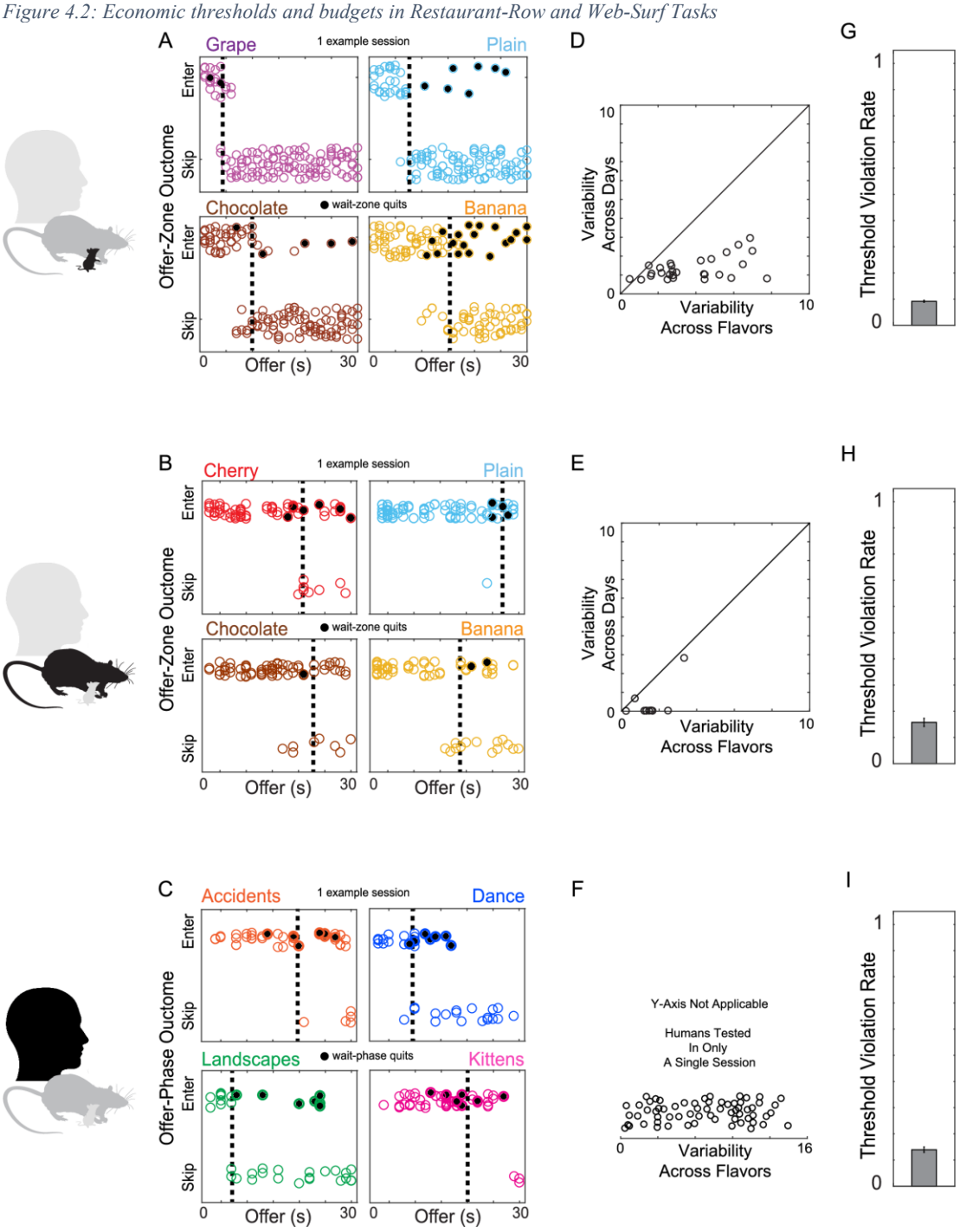
In the Restaurant Row and Web-Surf Tasks, decisions to abandon an on-going investment appear highly sub-optimal at face value, particularly if subjects were cued of the offer cost at trial onset before accepting. A potential optimal strategy could include making smart offer zone decisions to skip vs. enter informed by cued offer cost and never quitting once in the wait zone. Taking advantage of the economic nature of these tasks, we characterized the efficiency of observed behaviors by generating a computer model that predicted optimal number of rewards that could be earned if subjects were behaving as efficiently as possible. Optimal behavior was based on each individual's behavioral performance (e.g., reaction time and subjective threshold of willingness to wait in each restaurant) if they were behaving as efficiently as possible (using the fastest quartile of reaction and consumption times, by following their revealed preferences strictly, by never quitting). Proportion of actual observed earnings relative to model-predicted maximally optimal earnings was calculated.

Results

All three species learned to forage in a way that revealed preferences for certain rewards and all species used reliable subjective valuation strategies to decide between multiple competing reward offers (Figure 4.2).

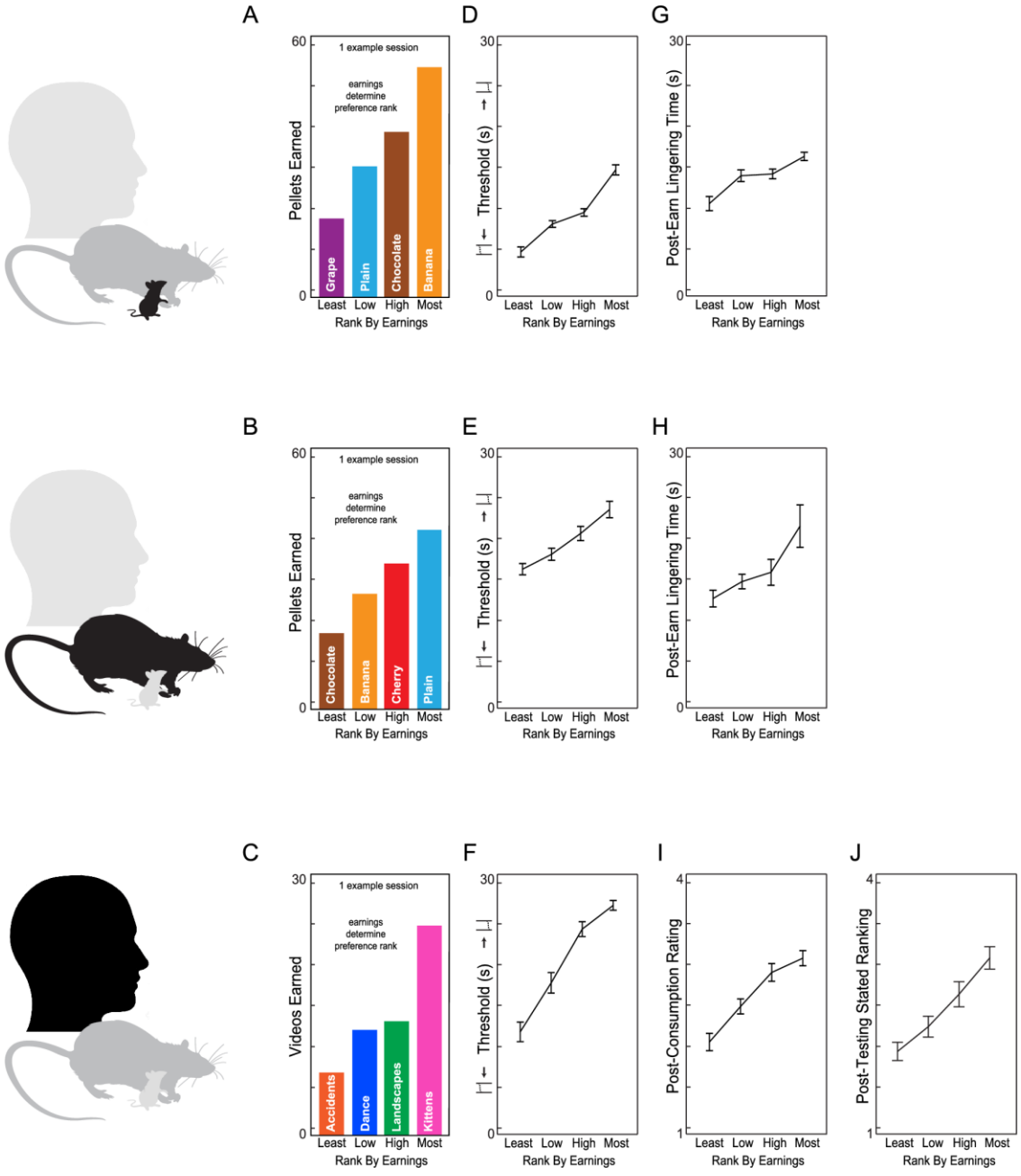
The value of a reward can be assessed in multiple ways. Often, this is measured in an instrumental manner, where how willing a subject is to take a reward (e.g., measured in amount of resources spent, effort expended, or behavioral invigoration) can reflect reward valuation. This is sometimes referred to as reward-seeking, reward-taking, or "wanting" valuations (Clark et al. 2012; Berridge 1996). In the Restaurant Row and Web-Surf Tasks, these are measurable in offer- and wait zone choice behaviors. These are distinct from post-consummatory valuations after an individual has earned a reward. Such post-consumption valuations are sometimes referred to as hedonic or "liking" valuations (Berridge 1996). In the Restaurant Row and Web-Surf Tasks, these can be measured after subjects earned rewards. In humans, this is more overtly assessed, since subjects were asked to rate each video on a scale from 1-4 (4=most enjoyed) immediately after viewing (or "consuming") the short (4s) video reward. These ratings matched other metrics of subjective valuation on this task (Figure 4.3C-J). Importantly, naturally enjoyable videos were used to mimic the same immediate consummatory nature of inter-trial pellet eating used in many appetitive-driven rodent laboratory studies (which is unlike many human studies that use hypothetical ratings, money, or other token-systems redeemed at the end of testing). Interestingly, we found that rodents, after earning and consuming food pellets on the Restaurant Row task, often lingered at the reward site before advancing to the next trial at the next restaurant (Figure 4.3G,H). Surprisingly, rodents typically spent ~50% of the entire 1hr testing session engaging in this lingering behavior. This decision to linger rather than leave, where no overt reward is being sought out, may represent a conditioned-place-preference-like effect associated with each restaurant's unique context (Clark et al. 2012). Interestingly, rodents lingered longer in more-preferred restaurants (Figure 4.3G,H). Therefore, we can take this behavioral metric as one that is (a) distinct from an instrumental reward-taking or "wanting" valuation, (b) appears to mimic post-consumption hedonic

Figure 4.2: Economic thresholds and budgets in Restaurant-Row and Web-Surf Tasks



(A-C) Mice (A), rats (B), and humans (C) entered low delays and skipped high delays in the offer zone, while infrequently quitting once in the wait zone (black dots). Dashed vertical black lines represent calculated offer zone and wait zone "thresholds" of willingness to budget time. Thresholds were measured from the inflection point of fitting a sigmoid curve to earns vs. not-earns as a function of offer cost. (D-F) Threshold variability between flavors along the x-axis in mice (D), rats (E), and humans (F) as well as within flavor across days along the y-axis in mice (D) and rats (E). Threshold variability across days in rodents was relatively stable. (G-I) Probability of trials that were either earned when the offer was greater than threshold or not earned when the offer was less than threshold in mice (G), rats (H) and humans (I).

Figure 4.3: Multiple valuation metrics of subjective preferences

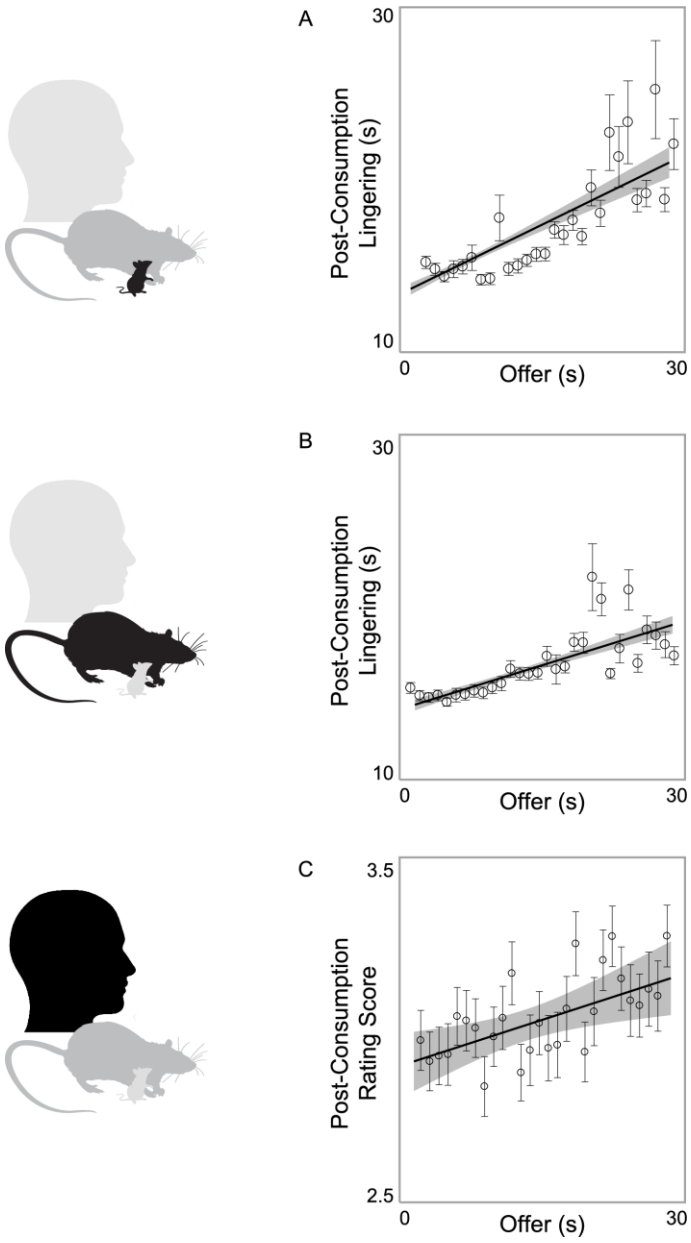


(A-C) Flavors were ranked from least- to most-preferred by summing the number of rewards earned in each restaurant or gallery in a single session. Panels show one example session in mice (A), rats (B), and humans (C). (D-F) Average thresholds sorted by rankings as defined in (A-C) were higher for more-preferred flavors in mice (D), rats (E), and humans (F). (G-I) Post-consummatory hedonic valuations sorted by rankings as defined in (A-C) we higher for more-preferred flavors in mice (G), rats (H), and humans (I). See Supplemental Discussion. (J) After the testing session was completed, human subjects stated rankings in a survey debrief session. Stated preferences were higher for more-preferred genres sorted by rankings as defined in (C). Multivariate Pearson and Spearman correlation analysis controlling for multiple comparisons found that all valuation metrics (earn rankings, calculated thresholds, post-consumption behaviors, and stated preferences [humans only]) significantly correlated with each other within each species (all correlations, $P<0.001$).

valuations measured in humans on the parallel Web-Surf Task, and (c) captures subjective flavor preferences that match other valuation metrics for unique flavors on this task, similar to humans (Figure 4.3).

Interestingly, we found that the degree of post-consumption hedonic valuations correlated with offer cost in mice, rats, and humans (Figure 4.4). That is, after subjects earned rewards that required more expensive investments, they hedonically valued those consumed rewards higher compared to earned rewards that were inexpensive. This is related to but distinct from the sunk-cost phenomenon, which is rooted in enhanced reward valuations. The sunk-cost phenomenon is classically described as an escalation of continued reward investment or commitment as a function of irrecoverable past investments made toward reward receipt. Thus, the sunk-cost effect is linked to enhancements in instrumental reward-taking valuations. The sunk-cost phenomenon, classically, makes no prediction with regard to the hedonic value of already owned or consumed rewards. While the two types of valuations are certainly related, they are separable on the Restaurant Row and Web-Surf Tasks. At the surface, because findings in Figure 4.4 are related to both post-earn hedonic valuations as well as instrumental reward-taking costs required to earn the reward itself, these data appear related to two other cognitive heuristics often compared to the sunk-cost phenomenon: the endowment effect and the post-purchase rationalization effect. The endowment effect posits that already owned objects are more highly valued than those not possessed (Kahneman et al. 1991). The post-purchase rationalization effect (sometimes described as "Buyer's Stockholm Syndrome" or "Marketing Placebo Effect") is a choice-supportive bias that posits more expensive rewards are more highly valued than less expensive rewards after purchasing them (Schmidt et al. 2017; Plassmann et al. 2008; Cohen and Goldberg 1970). This data is consistent with the post-purchase rationalization effect shared across species. While there have been reports of the endowment effect in non-human animals, to our knowledge this is the first report of evidence for the post-purchase rationalization effect in non-human animals (Lakshminaryanan et al. 2018; Smith and Smith 1982).

Figure 4.4: Offer cost and post-consumption valuations



(A-C) Hedonic valuation metrics measured immediately after consuming rewards as a function of offer length in mice (A), rats (B), and humans (C). Rodents (A-B) lingered at the reward site after consuming rewards before leaving for the next trial at the next restaurant (see Supplementary Discussion). Humans (C) rated videos on a scale from 1-4 (4=most enjoyed) immediately after viewing earned rewards. Post-consumption valuations positively correlated with offer cost in mice (A, Pearson coefficient $r = 0.737$, $P<0.001$), rats (B, $r = 0.733$, $P<0.001$), and humans (C, $r = 0.473$, $P<0.05$). (error bars ±1 SEM, shaded region represents 95% confidence interval of correlation).

The economic key to a foraging task is the division of time spent during the task. Our neuroeconomic task directly tests sensitivity to sunk costs across species using time as currency.

Flavors and genres allowed us to measure subjective preferences as a function of cost, avoiding the confound that different reward sizes might require different consumption times. The multiple zones allowed us to characterize multiple valuation processes involved in decisions: initial commitment valuations (offer zone), secondary re-evaluations (wait zone), and post-consumption hedonic valuations. In this task, two key factors minimized information uncertainty and automated reward-seeking behavior as potential confounds: (1) subjects were provided full information on cost and investment progress (tones counting down or download bar shrinking), and (2) earning rewards required subjects to wait and withhold quitting after making an initial acceptance decision (rather than requiring a repetitive action).

Mice, rats, and humans learned to forage economically on these tasks (Figure 4.2). We characterized multiple valuation metrics that revealed subjective flavor preferences similarly across dimensions in all species (Figure 4.3). We ranked flavors from least- to most-preferred by summing the number of rewards earned in each flavor at the end of every session (Figure 4.3A-C). We characterized how subjects budgeted their time within the limited session by calculating economic "thresholds" of willingness to earn rewards as a function of offer length in seconds (Figure 4.3D-F). Thresholds for different flavors varied across individuals in all three species (Figure 4.2D-F). Thresholds in rodents remained stable across many days for a given flavor for an individual (Figure 4.2D-E). Additionally, subjects rarely violated these thresholds (Figure 4.2G-I). We also characterized how subjects evaluated rewards post-consumption (Figure 4.3G-I). Lastly, in humans, we obtained stated preference rankings in a post-testing debrief survey (Figure 4.3J). All valuation parameters significantly correlated with each other in each species (Pearson and Spearman correlation coefficients all $P<0.001$, controlling for multiple comparisons).

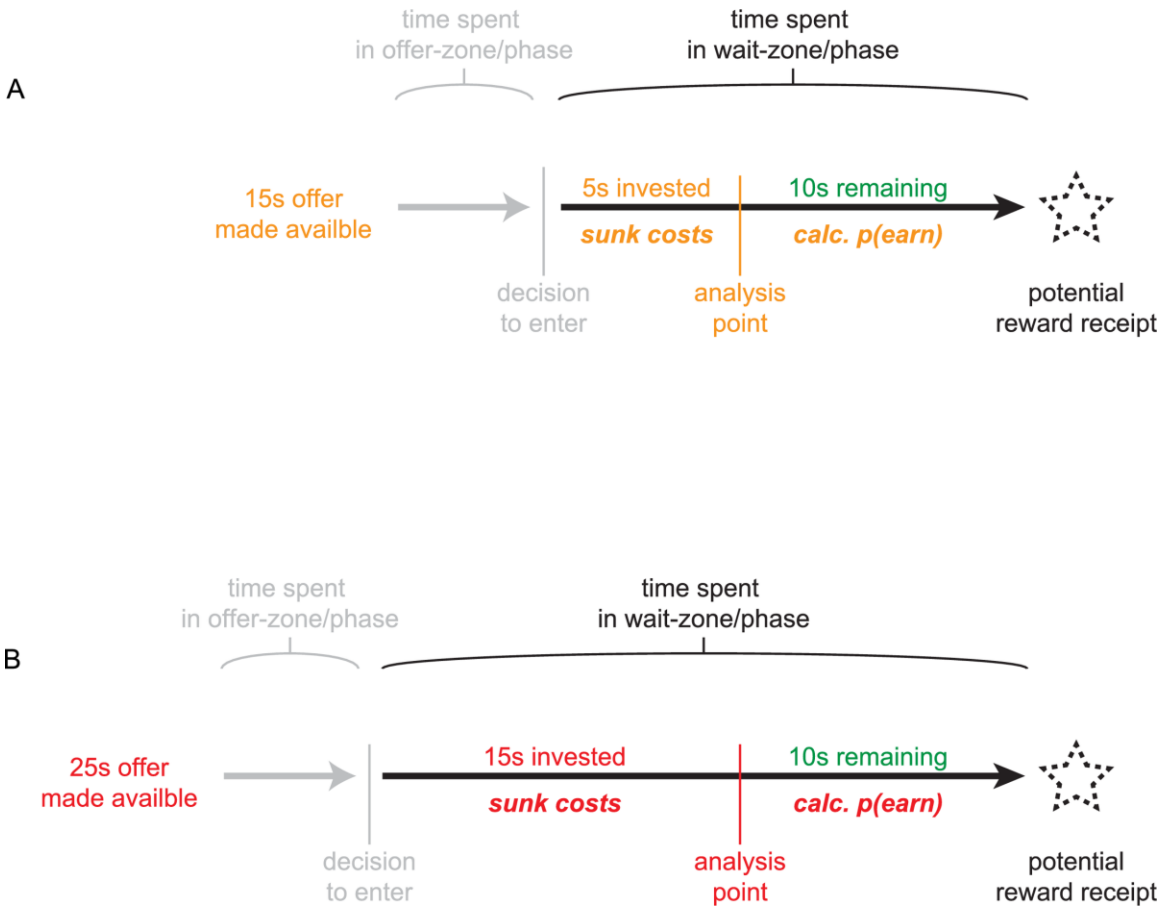
To address susceptibility to sunk costs, we examined quit decisions in the wait zone. These behaviors involve the abandonment of continued reward pursuit despite having made prior investments (partial waiting) while

on a limited budget (time). We parameterized the probability of earning a reward in the wait zone as a function of the remaining time-investment required to earn a reward (future costs) and the prior time-investment already spent waiting in the wait zone (past [sunk] costs, see Figure 4.5). The data yielded many samples across all conditions of time-remaining and time-spent (Figure 4.6), which allowed us to measure the extent to which irrecoverable prior investments (sunk costs) escalated wait zone commitment (Figure 4.7).

We found that mice, rats, and humans demonstrated robust sunk cost effects (Figure 4.8, ANOVA collapsing across all sunk cost conditions: mice: F=30.75, P<0.0001; rats: F=45.65, P<0.0001; humans: F=3.95, P<0.0001). Importantly, increasing prior investment amounts generated a continuously stronger sunk cost effect (Figure 4.8, example post-hoc comparison between +1s and +5s sunk costs: mice: F=45.49, P<0.0001; rats: F=54.41, P<0.0001; humans: F=4.21, P<0.05) – a critical tenet of the sunk cost fallacy (Arkes and Ayton 1999; Magalhães and White 2016).

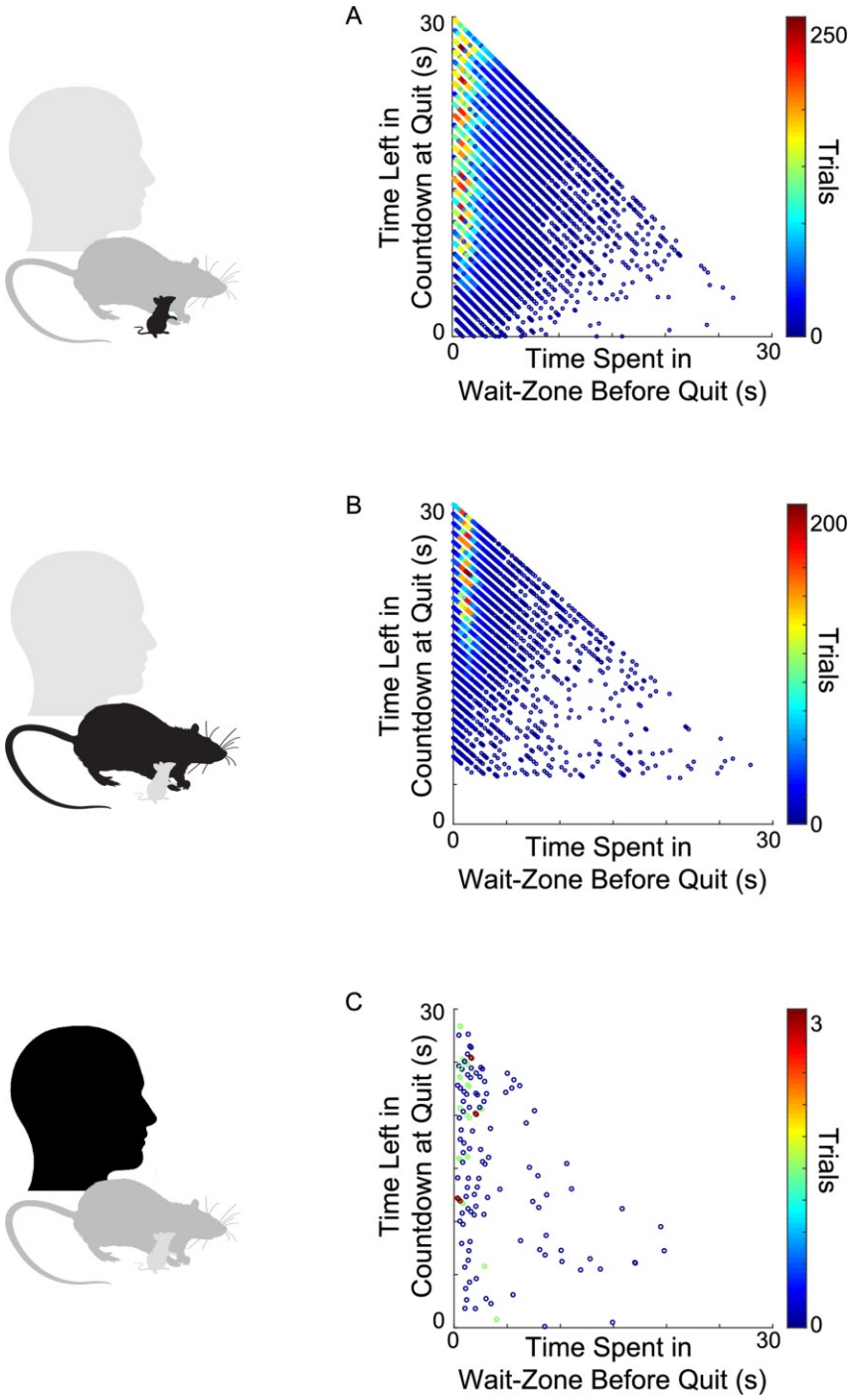
Time spent in the offer zone also detracts from the total time-budget, and a similar analysis can be performed (Figure 4.9). In contrast to re-evaluation processes in the wait zone, we found no effect of time spent in the offer zone. That is, the amount of time spent in the offer zone, surprisingly, did not alter the probability of earning rewards once in the wait zone (Figure 4.10, ANOVA collapsing across all offer conditions: mice: F=1.55, P=0.23; rats: F=0.77, P=0.39; humans: F=0.12, P=0.74). Importantly, the delay to reward did not start counting down while the subject remained in the offer zone. This meant that the animal was choosing between distant options and had not yet invested in the offer. This lack of an effect of time spent in the offer zone on progress abandonment once committed suggests that waste avoidance, overall resource depletion, and loss aversion are insufficient explanations of sunk cost-driven escalation of reward seeking behavior. This also suggests that the offer zone and wait zone may access separable valuation processes and reveals a novel determinant of susceptibility to sunk costs rooted in dissociable decision-making algorithms conserved across species.

Figure 4.5: Example economic scenarios in the wait zone sunk-cost analysis



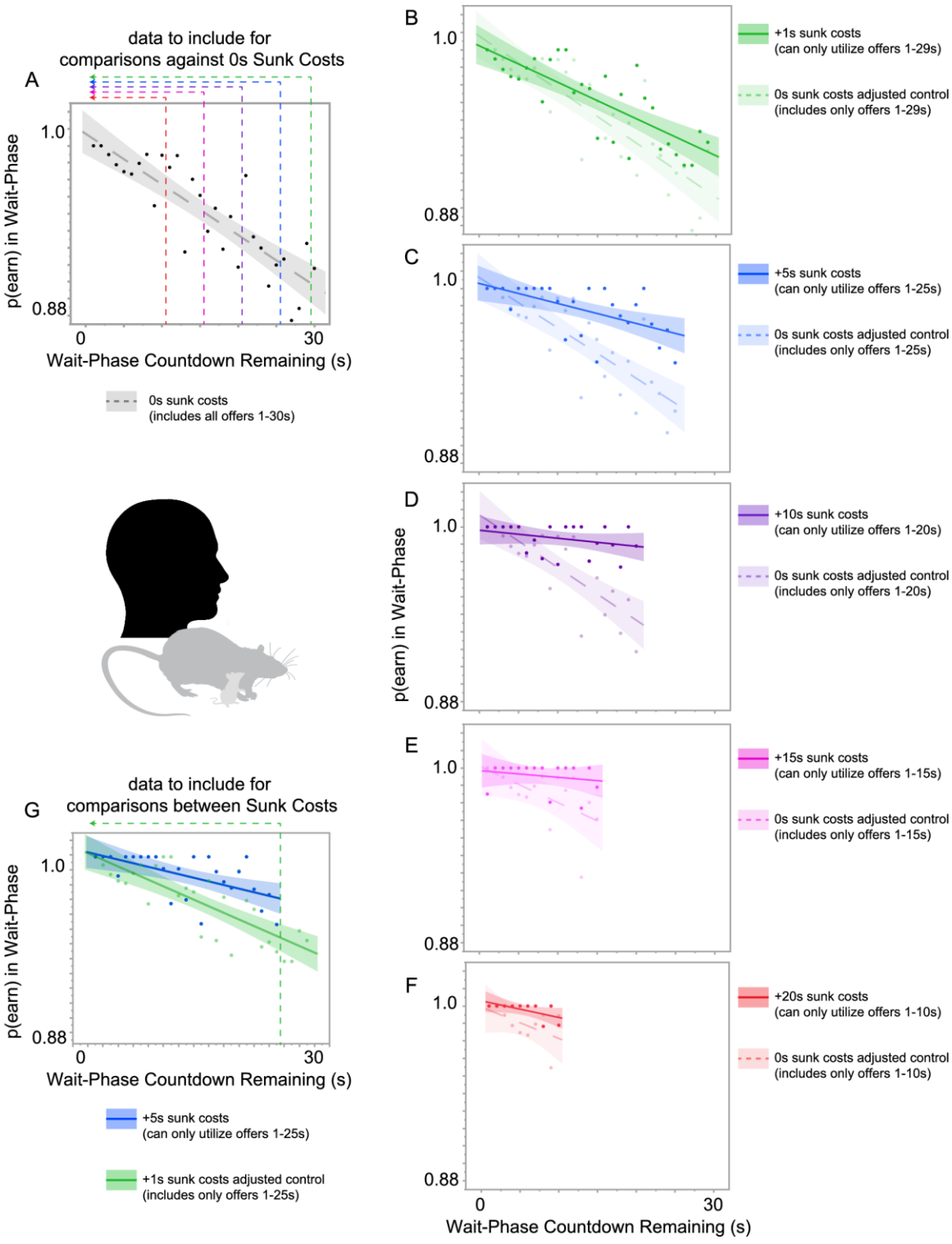
(A) Example in which a subject is offered a 15s-reward and chooses to enter the wait zone. [e.g., 100 cases of 15s offers entered]. Once in the wait zone, subjects could choose to quit at any moment. For example, taking trials where subjects had invested 5s in the wait zone and had not yet quit from that point onward [e.g., 25 trials already quit with <5s spent waiting, leaving 75 / 100 trials "surviving" so far] , we calculated the probability of earning a reward in the 10s window remaining in the countdown [e.g., 25 / 75 trials survived from that point forward until countdown completion = 0.33]. (B) Contrast with an example with a different initial starting offer: 25s. In this example, trials where subjects had invested 15s in the wait zone and had not yet quit have the same 10s window remaining in the countdown. We compared the p(earn) probabilities calculated in scenario A with the p(earn) probabilities calculated in scenario B. The sunk cost effect would predict an observed increase in p(earn) after investing 15s (B) compared to after investing 5s (A). The Restaurant Row and Web-Surf Tasks provide multiple initial starting delays of the offer as well as shifts in the analysis point, highly parameterizing the sunk cost effect along a continuum, along past-(sunk)- and future-cost dimensions.

Figure 4.6: Distribution of varying economic scenarios for use in the wait zone sunk cost analysis



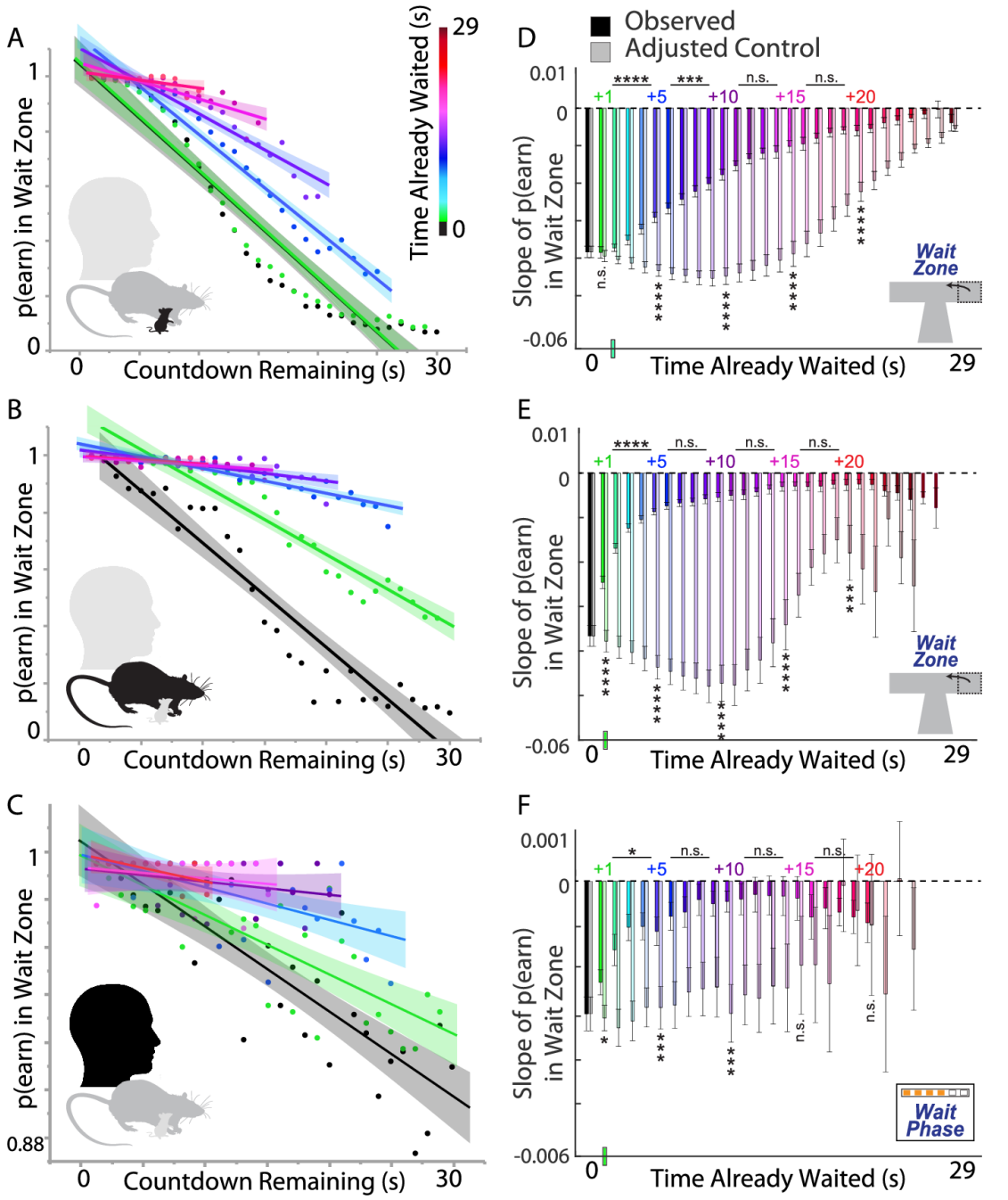
Each panel shows enter-then-quit trials plotting where time spent in the wait zone before quitting is plotted on the x-axis and time remaining in the countdown at the moment of quitting is plotted on the y-axis. Color axis indicates number of trials binned in each unique economic scenario for mice (A), rats (B), and humans (C).

Figure 4.7: Visualization of wait zone sunk cost analysis and controls



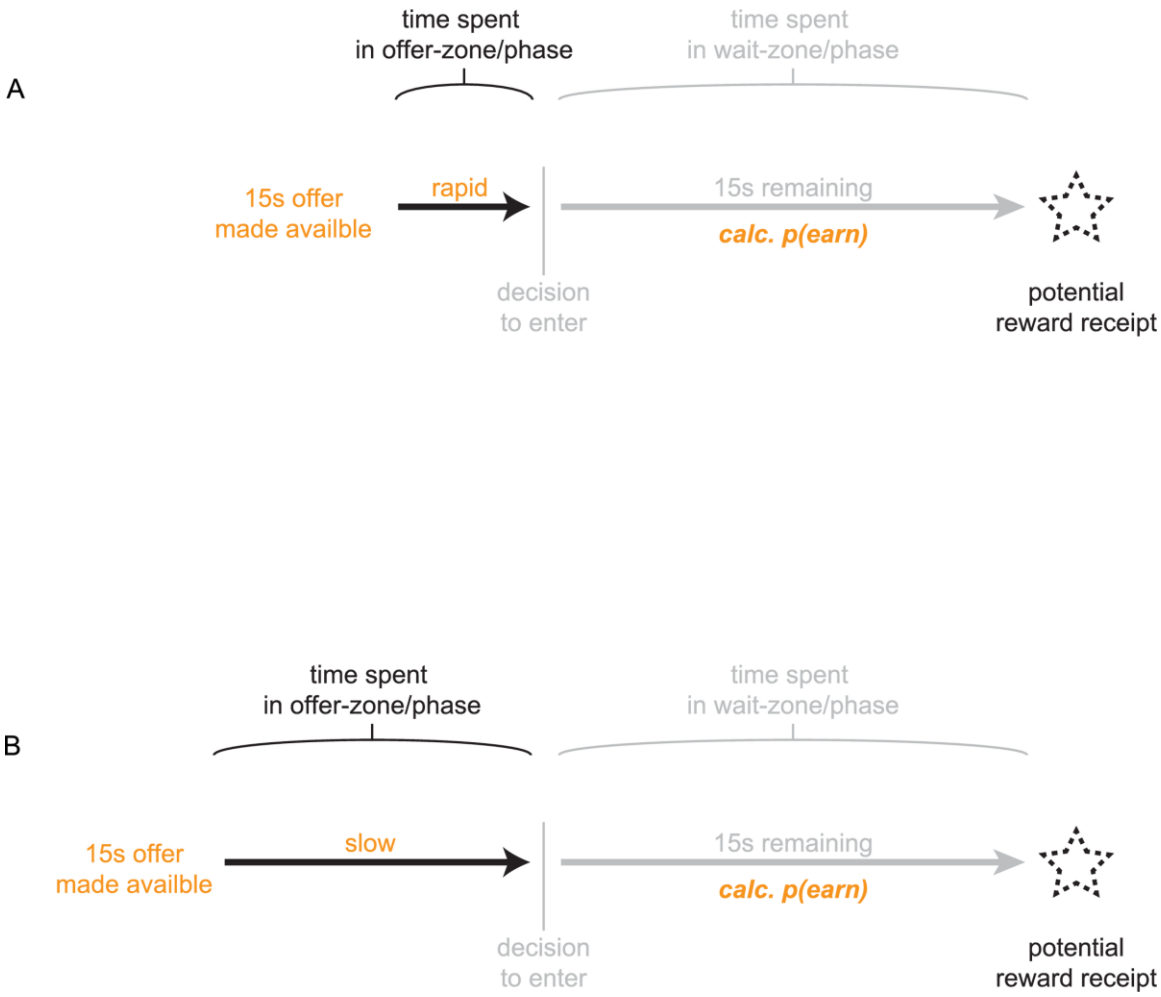
Using human data as an example, we ensured that our interpretation of slopes from linear regressions was not skewed by the different distributions of available data in the different sunk cost conditions. For instance, data from "30s remaining in the countdown" does not exist for the +1s sunk cost condition while it does exist for the zero sunk cost condition. To correct for this potential confound, we used regressions on the zero sunk cost condition iteratively leaving out the right most data points successively as the underlying control (A). Colored arrows indicate data that was included for each adjusted control regression: +1s-sunk costs (green), +5s-sunk costs (blue), +10s-sunk costs (purple), +15s-sunk costs (magenta), and +20s-sunk costs (red). (B-F) Examples of comparing regressions between each sunk cost condition and the zero sunk cost condition adjusted for the progressively smaller dataset illustrated in (A). (G) Similar concept of using regressions on progressively shortened datasets instead here to compare two sunk cost conditions against each other. In this example, the +5 sunk cost condition can only include offers from 1-25s. Thus, comparison against the +1 sunk cost condition, when adjusted, includes that same limited range of offers.

Figure 4.8: Time spent waiting increases commitment to continue reward pursuit cross-species



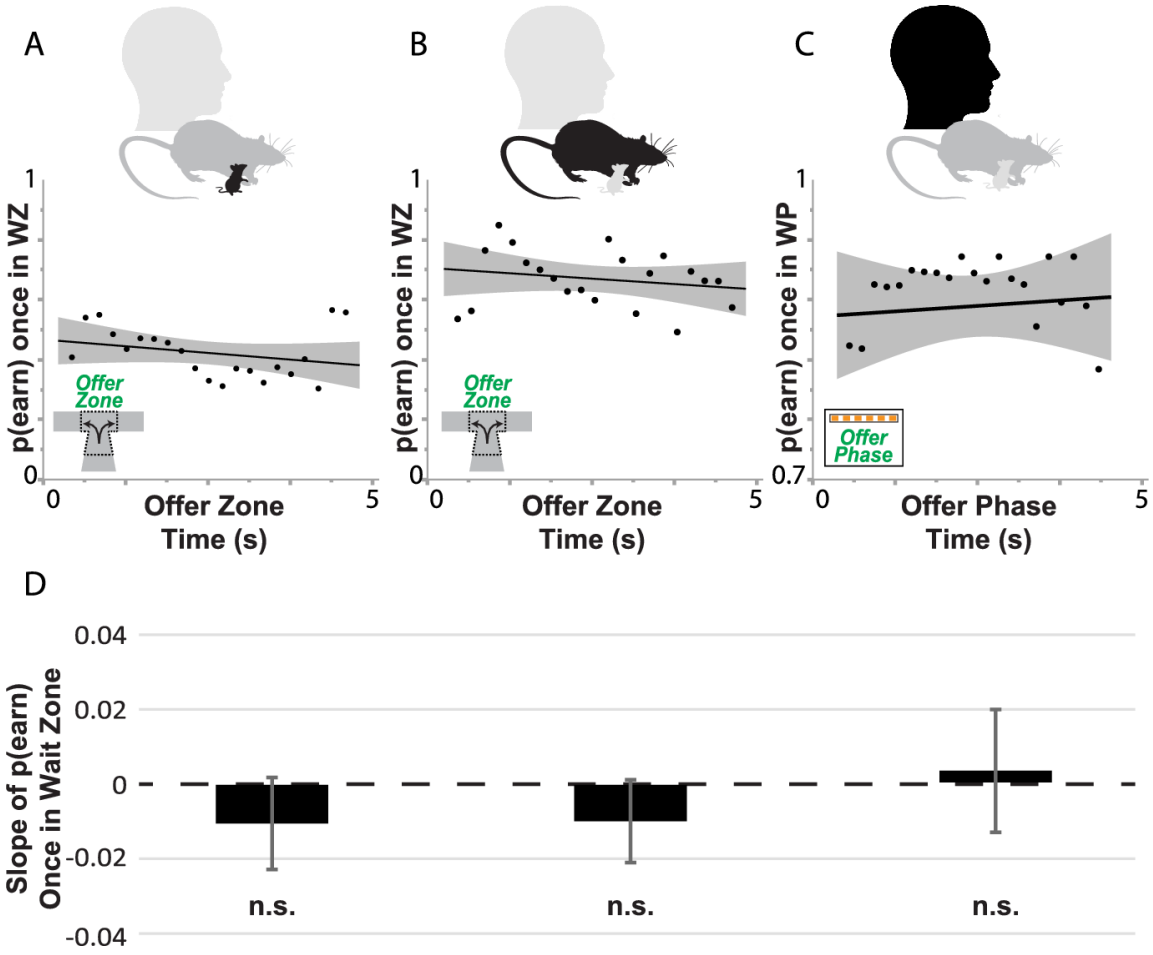
(A-C) Probability of earning a reward in the wait zone as a function of countdown time remaining: in (A) mice, (B) rats, and (C) humans. Black data points indicate trials where subjects had just entered the wait zone. Colored data points indicate time remaining in the countdown after subjects had already waited varying times (see Fig.S3). Linear regressions are plotted with 95% confidence interval shadings. (D-F) Slopes calculated from each linear regression in (A-C, "observed") are plotted ±1 SEM. Slopes re-calculated iteratively from black data points to match colored data ranges in (A-C, "adjusted controls," see Fig.S5). Colored tick on x-axis indicates time in wait zone until first significant sunk cost effect was observed. ANOVAs were used to compare slopes of linear regression models, testing for interactions with sunk cost conditions and controls, correcting for multiple comparisons. Not significant (n.s.) P>0.05, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Figure 4.9: Example economic scenarios in the offer zone sunk-cost analysis



In these two scenarios, subjects were offered a 15s-reward and chose to enter the wait zone. Scenario A illustrates a rapid decision to enter whereas scenario B illustrates a slow decision to enter where subjects invested more time in this initial offer zone decision. In both examples, the delay is the same (both 15s). We calculated the probability of earning once in the wait zone. The sunk cost effect, based on a resource-depletion/wastefulness-avoidance would predict an observed increase in p(earn) in (B) compared to (A). However, we found no such changes in mice, rats, or humans.

Figure 4.10: Resources spent while deliberating do not contribute to the sunk cost effect

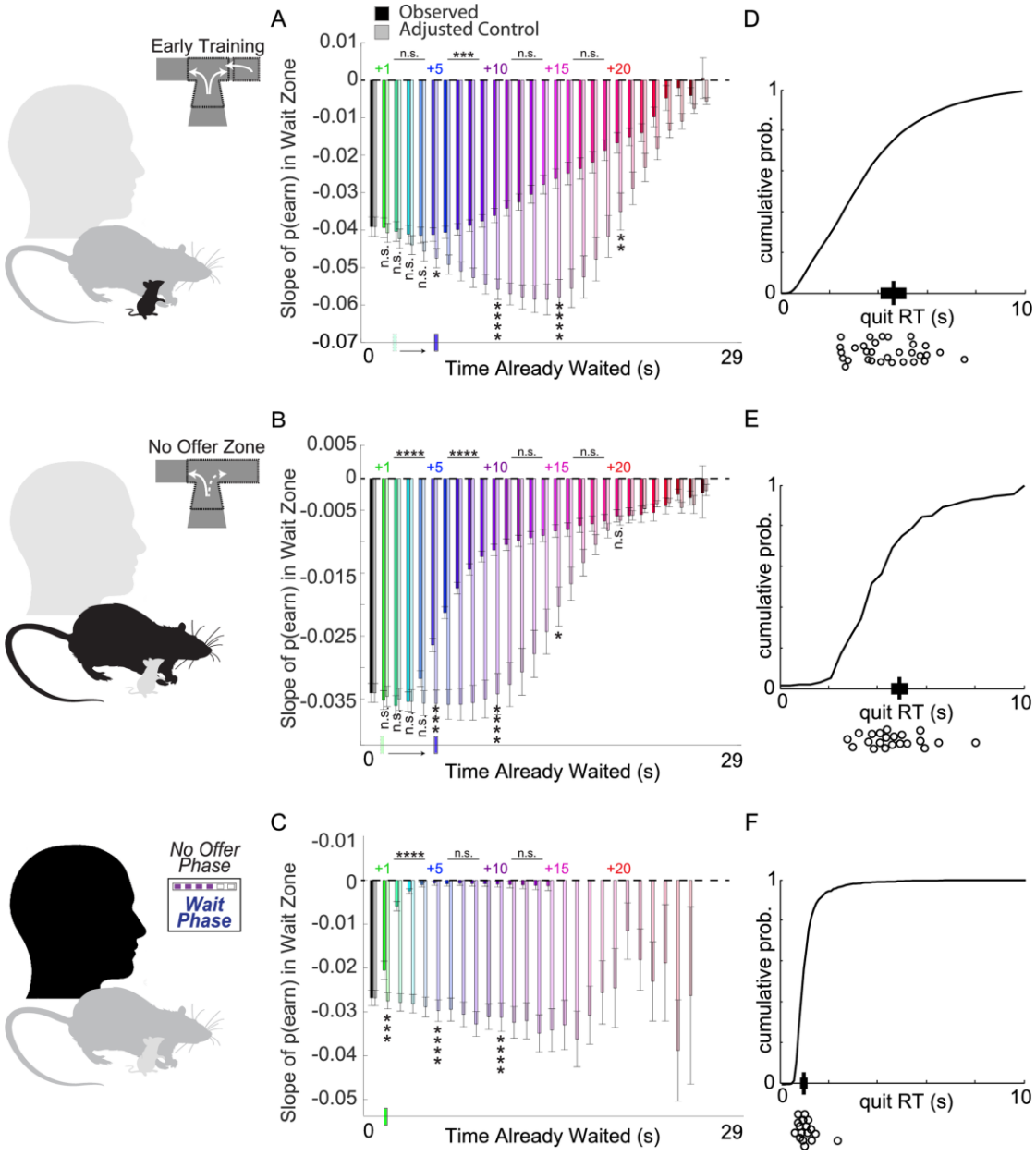


(A-C) Amount of time spent in the offer zone choosing to skip vs. enter did not influence the probability of earning vs. quitting once in the wait zone after subjects chose to enter (see Fig.S6). Linear regressions are plotted with 95% confidence interval shadings. (D) Slopes calculated from linear regressions are plotted ±1 SEM and are not significantly different from each other or zero in either mice ($F=1.545$, $P=0.229$), rats ($F=0.767$, $P=0.392$), or humans ($F=0.117$, $P=0.737$) using an ANOVA with post-hoc comparisons against zero. Not significant (n.s.) $P>0.05$.

In the offer zone, the reward delay does not start counting down toward reward delivery. Therefore, two critical differences from the wait zone may explain why sunk costs do not accrue in the offer zone: (1) time spent in the wait zone is valued differently because it is actually counted toward reward-earning progress; (2) countdown tones that descend in the wait zone (i.e., melodic contours) or a diminishing video download bar in the wait phase may carry added value learned through incentive salience (Graves et al. 2014; Clark et al. 2012). Both factors could weigh continued commitment in the wait zone disproportionately higher than quitting (i.e., enhance a loss-aversion bias) and could explain why sunk costs do not accrue in the offer zone.

To control for these two factors, we tested rodents and humans in variants of the Restaurant Row and Web-Surf Tasks in which the offer zone was not present (Figure 4.11). In these alternative task variants, time started counting down as soon as the subject entered a restaurant or video gallery. In rats, we found that the sunk cost effect (ANOVA collapsing across all sunk cost conditions, $F=33.93$, $P<0.0001$) did not begin to accrue until after an initial window of time had elapsed (Figure 4.11B, post-hoc tests: +1s: $F=0.08$, $P=0.78$; +2s: $F=0.22$, $P=0.64$; +3s: $F=0.01$, $P=0.93$; +4s: $F=2.15$, $P=0.14$; +5s: $F=12.69$, $P<0.001$). Despite detracting from the session's total time budget and despite counting as "down payments" progressing toward reward delivery, an initial portion of time spent was not counted toward sunk costs, mimicking offer zone findings. Thus, in a task variant without an offer zone, the sunk-cost effect in the wait zone had a delayed onset. This suggests a serial process of (1) choosing-between followed by (2) opting-out was still engaged, despite the lack of an explicit offer zone. In support of this, the window when sunk costs did not accrue matched the average reaction time it took rats to turn down offers (Figure 4.11E, ~5s). Humans on this variant of the task also displayed the sunk-cost effect (ANOVA collapsing across all sunk-cost conditions, $F=61.42$, $P<0.0001$), although a delayed onset was not detectable in these data (Figure 4.11C, post-hoc test: +1s: $F=12.29$, $P<0.001$, unlike rats on this variant). However, humans were significantly faster than rodents at turning down offers, and, consistent with rodents, the onset of sunk-cost effects matched this average reaction time (Figure 4.11F, ~1s). It is possible that humans similarly progressed through a serial process of choosing-between followed by opting-out but at a faster pace than rodents.

Figure 4.11: Additional sunk cost analyses across training and in other task variants.

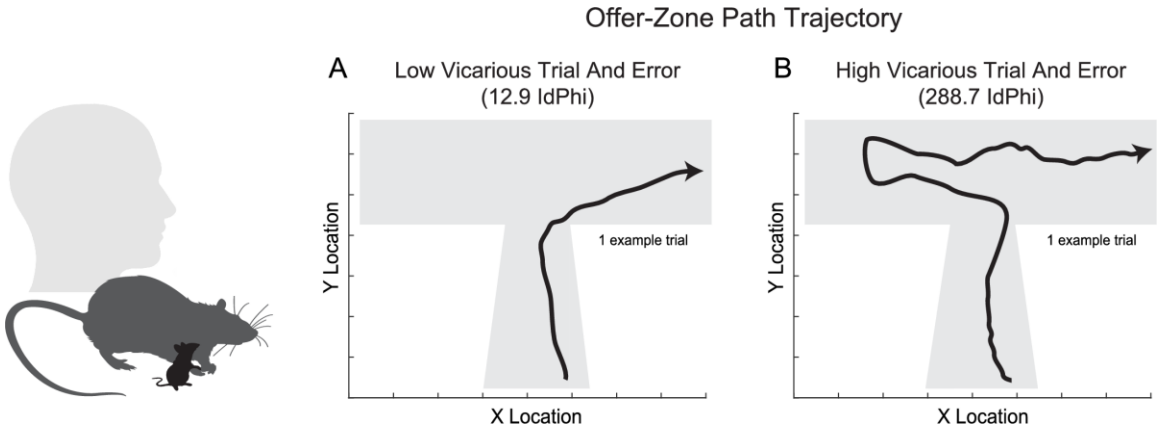


(A) Mice early on in training still displayed sunk cost effects, however did so with a delayed onset compared to late in training (see Fig.2D). Blue colored tick on x-axis indicates time in wait zone until first significant sunk-cost effect was observed. Transparent green tick replotted from Fig.2D for reference in mice with extensive training. (B-C) A separate cohort of rats (B) and humans (C) were trained on variants of the Restaurant Row and Web-Surf Tasks without an offer zone where countdown began immediately and subjects only made quit decisions. Colored tick on x-axis indicates time in wait zone until first significant sunk-cost effect was observed. (B) Transparent green tick replotted from Fig.2E for reference in rats on the task variant with an offer zone. (D-F) Cumulative probability distribution of quit reaction time in the wait zone in mice (D), rats (E), and humans (F). Black tick on x-axis indicates cohort average reaction time ±1 SEM. Dot scatter below axis represent individual subjects' average reaction time. ANOVAs were used to compare slopes of linear regression models, testing for interactions with sunk cost conditions and controls, correcting for multiple comparisons. Not-significant (n.s.) P>0.05, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

In the offer zone, rodents displayed “pause and look” behaviors, known as “vicarious trial and error” (Figure 4.12, Tolman 1939; Redish 2016; Muenzinger 1956). Vicarious trial and error behaviors reveal on-going deliberation and planning during moments of indecision (Redish 2016; Muenzinger 1956; Tolman 1939). Numerous in vivo electrophysiology recording studies have demonstrated that during these behaviors, hippocampal representations sweep forward along the path of the animal, alternating between potential goals until the animal knows where to go (Johnson and Redish 2007; Redish 2016). Such goal representations are synchronized to reward-value representations in the ventral striatum, suggesting outcome predictions are being evaluated serially (van der Meer and Redish 2009; Stott and Redish 2014).

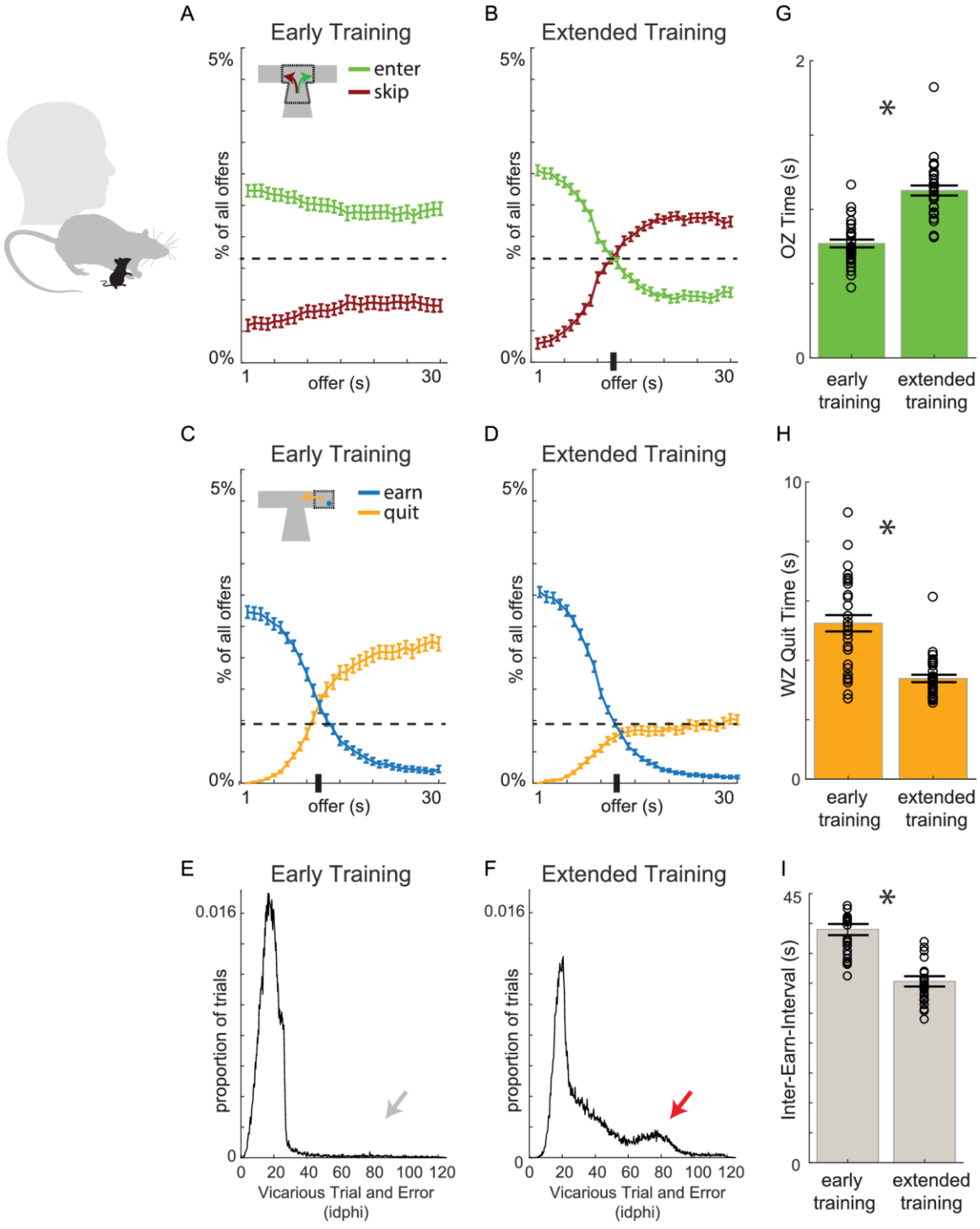
Restaurant Row training in mice (offer zone present) provided further insight into the development of a serial process decision stream: initially deliberating followed by re-evaluations to opt-out. Mice were tested daily up to 115 days, far longer than rats (typically run for 40 days) and humans (tested in a single session). We observed pronounced changes in decision strategies between early and late training as mice were still learning the structure of the task. Early in training, mice rapidly accepted all offers indiscriminately without factoring in the randomly selected 1-30s offer cost, and thus relied on quits to turn down expensive offers (Figure 4.13). In other words, offer zone enter decisions early in training were automated. This was likely a learned behavior in even earlier stages of training when offers were relatively inexpensive (see Methods, progressive stages: 1s offers only, offers ranging between 1-5s, and then between 1-15s), when cost information could be discarded, and when mice could afford to accept all offers. Interestingly, on these early 1-30s training sessions, we found that the sunk cost effect (Figure 4.11A, ANOVA collapsing across all sunk cost conditions, $F=4.46$, $P<0.0001$) did not begin to accrue in the wait zone until an initial window had elapsed (post-hoc tests: +1s: $F=0.14$, $P=0.71$; +2s: $F=0.28$, $P=0.60$; +3s: $F=0.65$, $P=0.43$; +4s: $F=1.67$, $P=0.20$; +5s: $F=4.10$, $P<0.05$) that matched the average reaction time it took mice to quit offers (Figure 4.11D, ~5s). This is reminiscent of the delayed onset of sunk costs observed in rats on the task without an offer zone when vicarious trial and error behaviors typically occurred (compare to Figure 4.11B,E). That is, despite separating each restaurant into an offer zone and wait zone, mice early in 1-30s training essentially ignored the offer zone. Importantly, mice did not display vicarious trial and error behaviors in the offer

Figure 4.12: Measurement of offer zone vicarious trial and error (VTE) in rodents



Illustrated here are X-Y-locations of a rodent's path-trajectory in the offer zone over time during two example trials. Vicarious trial and error measures the absolute integrated angular velocity over time and is measured in IdPhi units. (A) Example of a low vicarious trial and error event without any re-orientations. (B) Example of a high vicarious trial and error event with re-orientations at the choice-point.

Figure 4.13: Change in decision-making behavior over training in mice on the Restaurant Row task



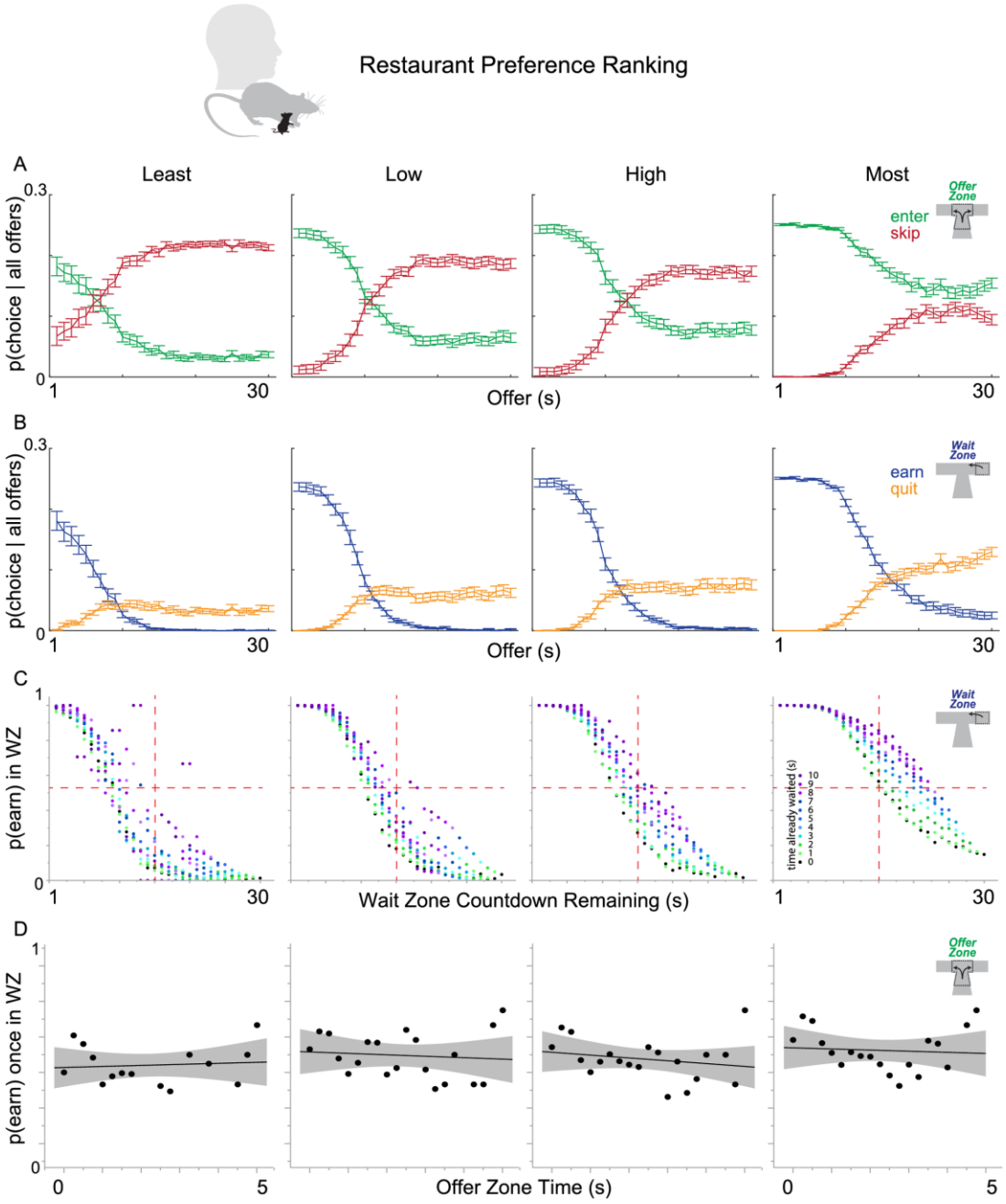
(A-B) Offer zone choice probability between skipping vs. entering as a function of offer length early in training (see Supplementary Discussion). (A) and after extended training (B). Mice entered nearly all offers indiscriminately early in training (A) while later learning to skip expensive offers (B). (C-D) Wait zone choice probability between quitting vs. earning as a function of offer length early in training (C) and after extended training (D). Mice learned to quit less with extended training, as they were more likely to accept offers in the offer zone they would be willing to earn. Black tick on the x-axis in B-D indicate cross over point of skip/enter or quit/earn decisions. Horizontal dashed black line indicates choice probabilities if decisions were random. (E-F) Probability density plots of vicarious trial and error behavior in the offer zone across early (E) vs. extended training (F). Red arrow in F emphasizes the presence of high vicarious trial and error events as mice learned to deliberate in the offer zone between cheap and expensive offers, illustrated in B. Gray arrow in (E) emphasis the absence of these high vicarious trial and error events as mice made snap judgments to accept all offers indiscriminately, illustrated in A. (G-I) As a consequence of learned changes in decision-making strategies over training, offer zone reaction time increased (G, $F=67.38$, $*p<0.0001$), wait zone quit time decreased (H, $F=36.48$, $*p<0.0001$), and reinforcement rate increased (I, decrease in inter-earn-interval time between pellet consumption, $F=45.26$, $*p<0.0001$). (open circles in (G-I) represent individual animals, error bars ±1 SEM)

zone early in 1-30s training (Figure 4.13E-F). This suggests that an initial deliberation decision was not actually made in the offer zone and that a true initial choose-between decision was not made until after mice entered the wait zone. Therefore, this would delay the onset of a secondary opt-out process in the wait zone. Taken together inexperienced mice transitioned between deliberation and foraging decision modalities once in the wait zone while experienced mice separated this transition between the offer- and wait zones. Importantly, the sunk cost effect tracked this transition across training.

Well-trained mice were fully capable of discriminating randomly presented cued offer costs in the offer zone differently in the different restaurants, ensuring information uncertainty was not a confounding factor in rodents (Figure 4.14). Furthermore, sunk cost effects occurred in all restaurants regardless of subjective flavor preference ranking, again only in the wait zone and not offer zone (Figure 4.14).

We ran additional analyses in order to control for possible differences in subjective value when entering the wait zone on different trials of varying offer lengths, even within the same restaurant, that may contribute to additional aspects of valuation that may confound interpretations of the sunk cost analysis. For instance, when entering the wait zone during a 25s offer for chocolate vs. entering the wait zone during a 15s offer for the same flavor, it is possible that those trials select from instances where the subject value of chocolate, independent of investment history, on that specific trial are biased toward high subjective valuations. That is, animals might just be "craving" chocolate more on those trials in which a 25s offer was entered compared to a 15s offer that was entered. Therefore, in our sunk cost analysis, it is possible that given a particular analysis time point (e.g., time left both being equal to 5s remaining in the countdown), the trials that started at 25s (with a time already waited equal to 20s) simply have a higher valuation of chocolate compared to the trials that started at 15s (with a time already waited equal to 10s), and thus the seemingly apparent effects of "sunk costs" on escalation of commitment in fact is higher simply due to trial selection biases and not actually differences in investment history.

Figure 4.14: Asymmetries in choices split by subjective value reassures cost discriminability in mice



Vertical columns split data by rankings (categorized in Fig.S2). (A-B) Choice outcome probabilities in the offer zone (A) or wait zone (B) relative to all offers encountered across all restaurants for a given offer length. Because offers are randomly presented from trial to trial, differences in offer zone ability to skip vs. enter high cost offers separated by subjective value helps ensure mice are capable of discriminating tones and information uncertainty is less of a confound. (C) Sunk cost effects in the wait zone exist in all restaurants (least: $F=3.33$, $p<0.05$; low: $F=4.34$, $p<0.05$; high: $F=10.11$, $p<0.01$; most: $F=26.69$, $p<0.001$). (D) Sunk cost effects in the offer zone do not exist in all restaurants (Pearson coefficient: least: $r=0.089$, $p=0.75$; low: $r=-0.097$, $p=0.69$; high: $r=-0.207$, $p=0.38$; most: $r=-0.077$, $p=0.74$). p(earn) in (C-D) are relative to each restaurant's offers. (error bars ±1 SEM, shaded region represents 95% confidence interval of correlation)

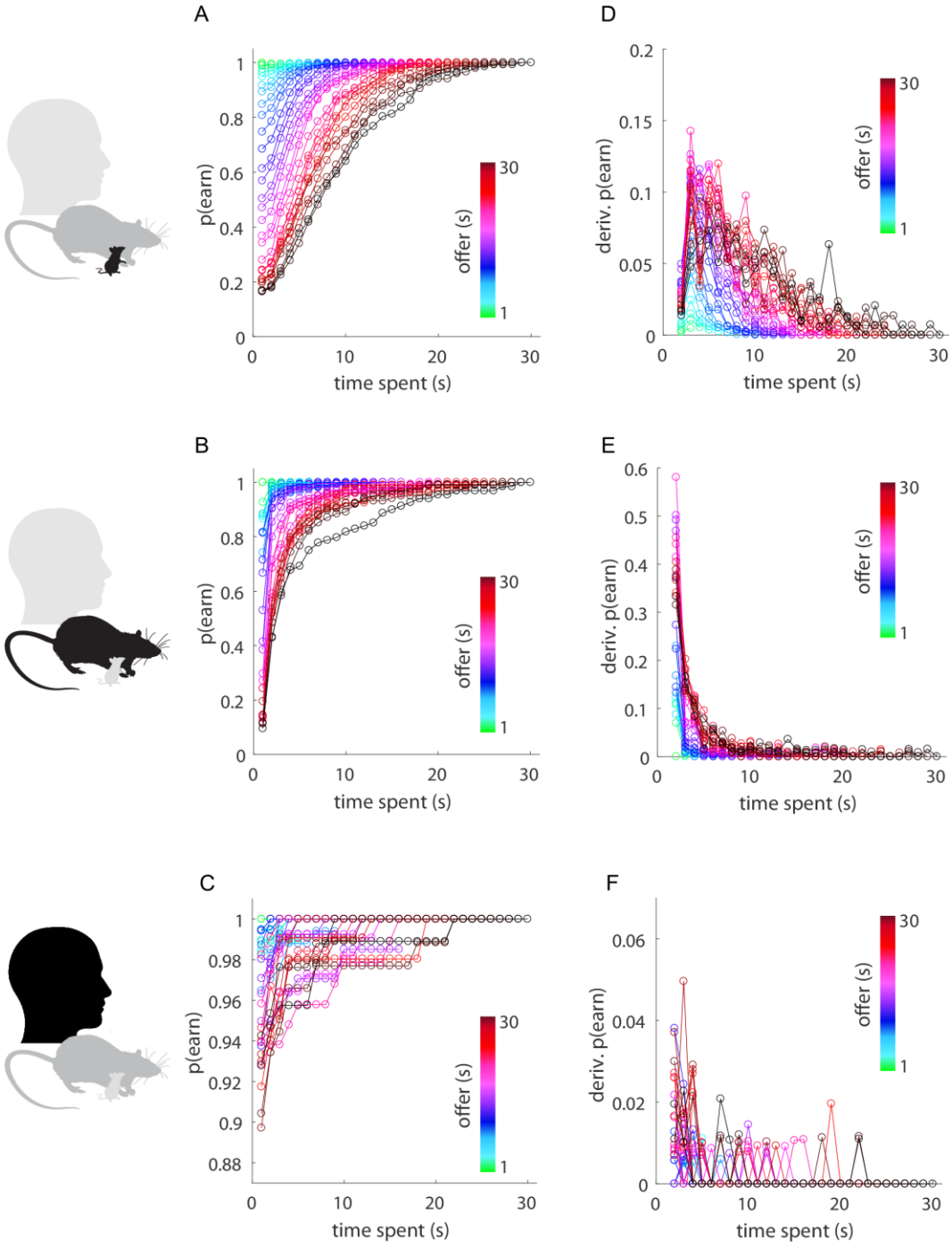
To directly test this concern, we ran additional analyses re-sorting our main analysis of wait zone sunk costs by splitting trials by starting offer length and calculating the probability of earning a reward as a function of time spent (Figure 4.15, as opposed to the original analyses that calculate p(earn) as a function of time left split by time already waited).

If the above concern was truly driving the apparent "sunk cost" effect, then if investment history played no role inflating reward value and escalating commitment in the wait zone, p(earn) should increase linearly as a function of time spent. That is, if 25s trials simply start off at a higher value than a 15s offer, what takes place during the countdown period should have no additive effect on escalating p(earn) as subjects continue to wait in the wait zone.

We found that the relationship between p(earn) and time spent waiting for each offer length trial type did not increase linearly with time spent but rather accelerated as a function of time spent (Figure 4.15). This suggests that investment history in the wait zone, as a function of time spent, carries added value that is not fully explainable by subjective value trial selection biases.

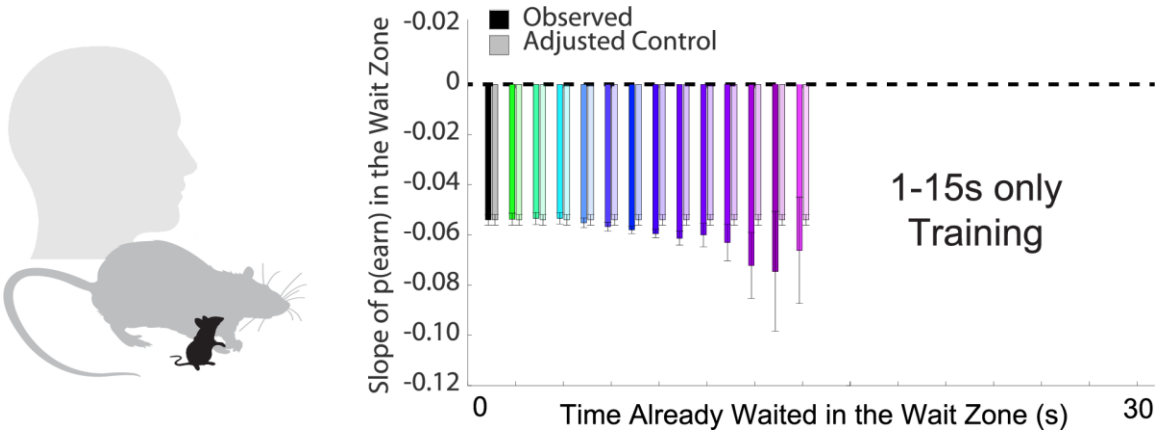
The sunk-cost effect may also relate to the amount of reward available in the environment. In a reward-scarce environment, the importance of decision optimality is higher than in a reward-rich environment. In other foraging tasks, rodents have been shown to adopt sub-optimal foraging strategies exacerbated in reward-scarce environments (Wikenheiser et al. 2013). These reports hint at sunk-cost-like effects that depend on the subjective over-valuation of time invested as a function of the state of the environment. Interestingly, mice training in relatively reward-rich environments (where offers ranged from 1-15s) did not demonstrate sunk-cost effects (Figure 4.16, ANOVA collapsing across all sunk-cost conditions, $F=1.13$, $P=0.35$), consistent with these reports. This indicates another important determinant of the sunk-cost phenomenon – environmental reward scarcity – that may have been overlooked in past studies presenting conflicting findings across species. This data is consistent with notions in other work that demonstrate State-Dependent Valuation Learning (SDVL) depends on the subject being in a leaner state and not

Figure 4.15: Sunk cost analysis re-sorted by offer length and time spent



(A-C) displays the original sunk cost analysis re-sorted to show the probability of earning a reward in the wait zone as a function of time spent waiting split by the original starting delay of the offer that was accepted when mice entered the wait zone in mice (A), rats (B), and humans (C) on the main version of the tasks with separate offer and wait zones. (D-F) displays the derivative of (A-C), illustrating the non-linear acceleration of p(earn) as a function of time spent.

Figure 4.16: Wait zone sunk cost analysis in mice in a reward-rich environment



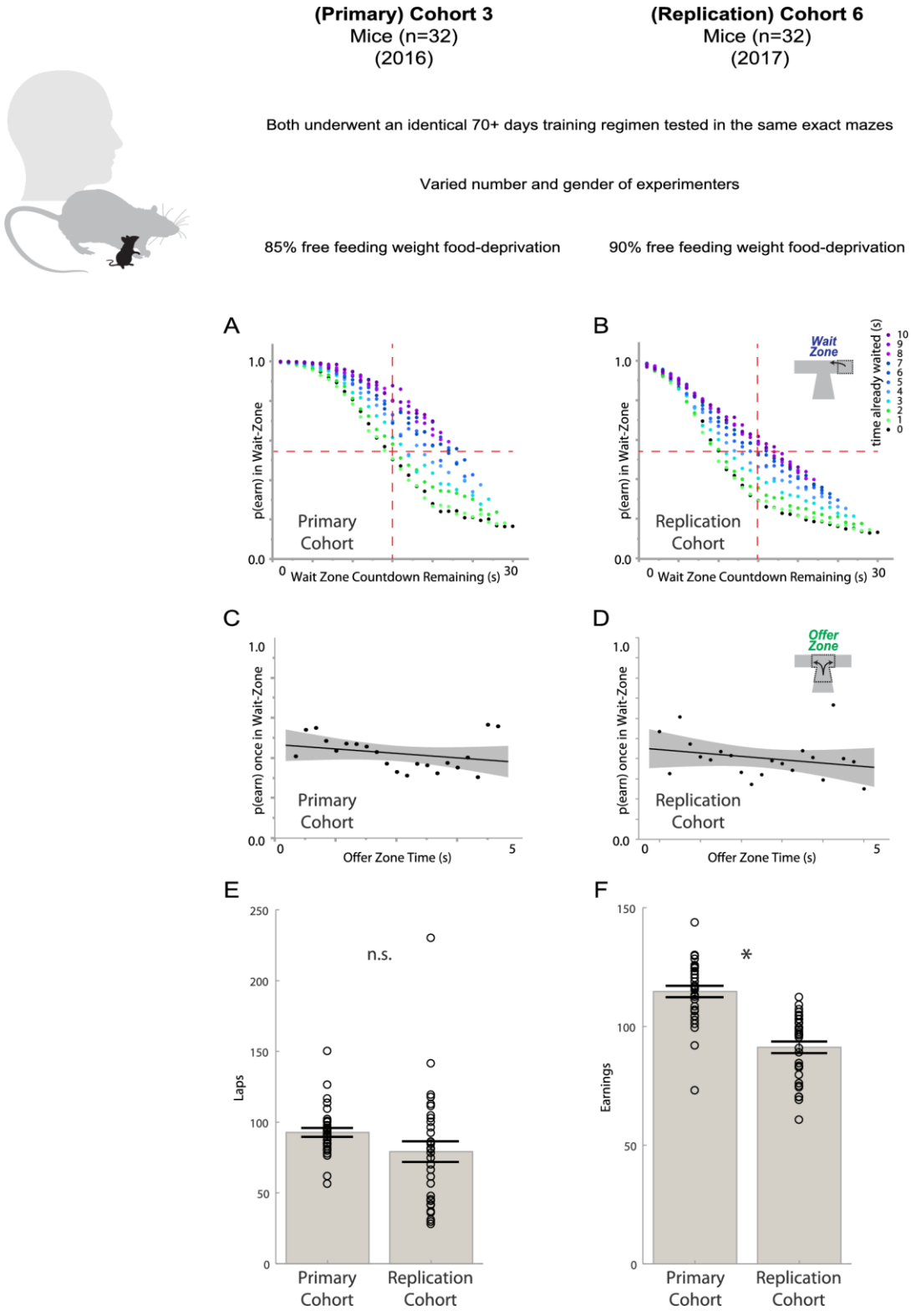
Before being exposed to the full range of 1-30s offers in the Restaurant Row task, mice were trained on 1-15s offers where reward cost was relatively inexpensive (see Methods, supplementary text S6). Slope of linear regressions plotted with ±1 SEM. There were no significant differences (P>0.05) between sunk costs conditions, controlling for adjusted data distributions previously described.

necessarily on lack of training (Pompilio et al. 2006; Aw et al. 2011). This helps explain why studies of Within-Trial Contrast (WTC) may have had unreliable outcomes. We explain both of these concepts further below in the discussion. Briefly, SDVL and WTC explain sunk cost-driven escalation of commitment following invested work occurring from depletion in physiological or psychological energetic states, thus making the yet-obtained reward seem more valuable in comparison to self.

If sunk-cost effects existed all the time in a leaner environment in our tasks (in the offer zone and wait zone, or during all moments in the wait zone in the task variants without an offer zone), sunk-cost effects on our task could be fully explainable by SDVL. The fact that decision processes seem to be interrupted or segregated into distinct stages within the same trial only in leaner environments and that these separate stages are differentially susceptible to sunk costs reveals dissociable valuation algorithms are at play within the same trial. Furthermore, aligning with the rationale posited in other work in response to reasoning for lack of WTC effects due to lack of sufficient training in other studies, effects reported in Figure 4.16 (lack of sunk cost effects even in the wait zone after any duration of waiting) occurred after 17 days of consecutive training (Aw et al. 2011). This included exposure to roughly 4,500 trials overall on average per mouse by day 17, or 1,500 trials of 1-15s offers in that training block (days 13-17), with 100 trials per newly experienced offer stimulus in that block (tones representing 6-15s offers). The previous block of training (days 8-12) included 1-5s offers and thus added additional exposure (roughly 400 trials) to 5s tones (capable of eliciting sunk cost effects in leaner 1-30s environment), by the 1-15s offer block. Despite any of this, tones only provide extra information that SDVL and WTC, in theory, could function without in order to drive sunk cost effects in this task design.

Other environmental factors including different experimenters or severity of food restriction did not impact main sunk cost findings in a replication cohort of mice that were tested by different experiments and food deprived to a lesser degree, despite influencing overall devalued economic behaviors likely due to satiety effects (Figure 4.17).

Figure 4.17: Replication cohort of mice varying environmental factors



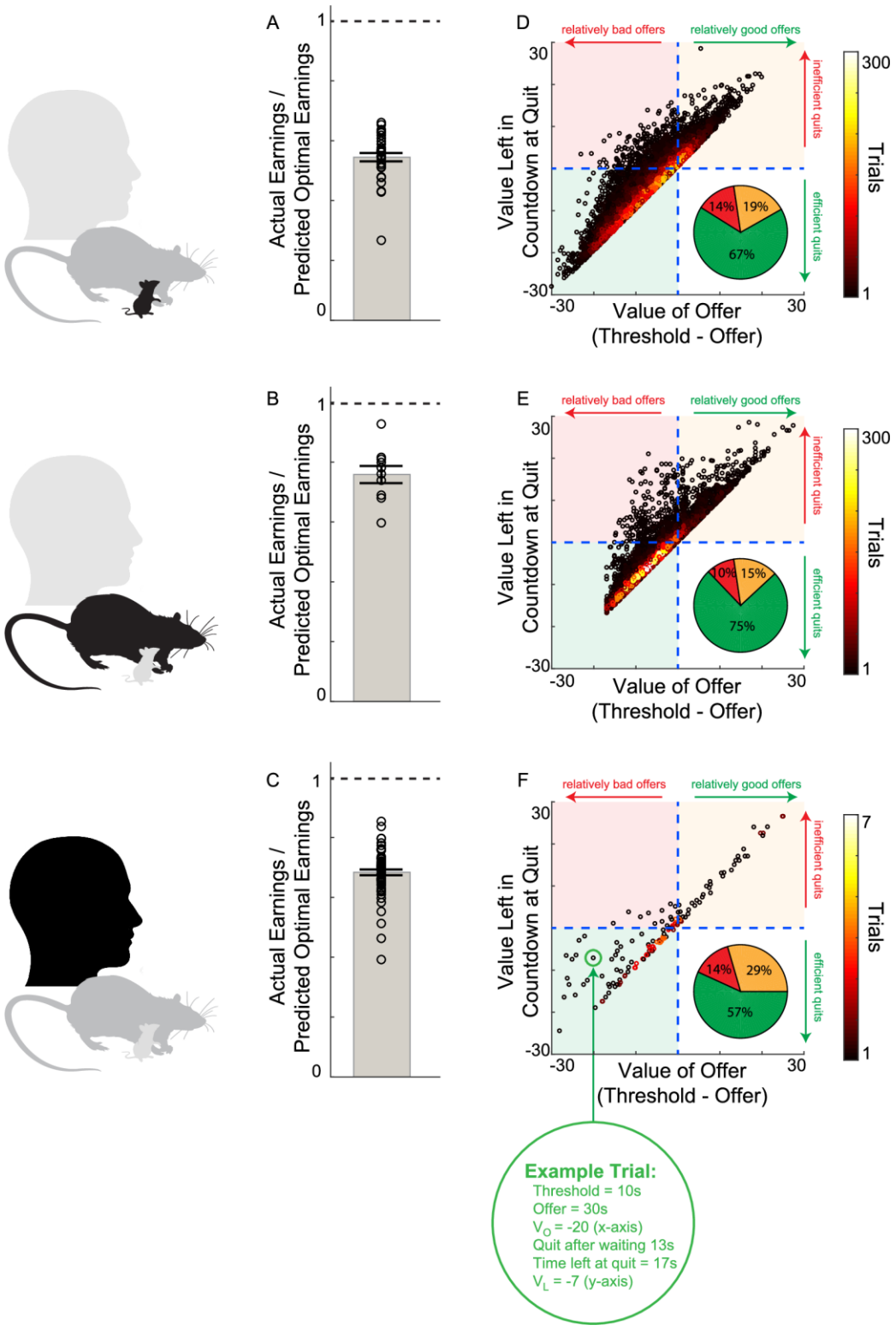
An additional 32 mice were tested on the main variant of the task (offer zone and wait zone). Experimenters were varied, and these mice were trained in the exact same apparatuses as the original cohort under the exact same training protocol. Additionally, these mice were intentionally food deprived to a lesser extreme. Despite these different environmental factors, mice still demonstrated a robust sunk cost effect in the wait zone (B, $F=26.56$, $p<0.01$, see original cohort in A for comparison) but no effect as a function of time spent in the offer zone (D, $r=-0.254$, $p=0.27$, see C for comparison). Environmental factors, while not affecting locomotor abilities nor number of offers encountered by not altering number of laps run (E, $F=2.86$, $p=0.10$), did result in a reduction in average number of pellets earned on the task (F, $F=46.79$, $*p<0.0001$). This is reflected in an overall left shift of the economic budget curves in panel B (vertical red dashed line indicates 15s, horizontal red line indicates 0.5 p(earn)), reflecting an alteration in reward value and willingness to wait. (open circles in (E-F) represent individual animals, error bars ±1 SEM, shaded region represents 95% confidence interval of correlation, not significant, n.s.)

To determine whether behavior on these tasks was sub-optimal, we used a computer-simulation model to predict the maximal rewards subjects could earn if they behaved optimally by following their revealed preferences strictly by not abandoning accepted offers in the wait zone (see Methods). We found mice, rats, and humans indeed performed sub-optimally (Figure 4.18A-C). This sub optimality arose from disadvantageous wait-zone decisions (Figure 4.18D-F). We can use an individual's restaurant's threshold to calculate the relative value of each offer encountered. The value of the offer (VO=Threshold – Offer) plotted against the value of the time left in the countdown for a given trial at the moment of quitting (VL=Threshold – Time Left To Go) reveals the economic efficiency of wait zone decisions. Economically disadvantageous wait-zone quit decisions drive sub optimality through a susceptibility to sunk costs.

While the primary sunk cost analyses used in this chapter measured the effects of time already invested on escalating the probability of earning a reward in the wait zone, the Restaurant Row task allows us to operationalize the subjective value of a given offer. The thresholds calculated for each subject are stable within or between sessions and allow for a way to normalize the value of each offer. That is, if a subject's threshold for a given restaurant is 15s and an example offer on a particular trial is also 15s, one could calculate the value of the offer by subtracting the offer length from the threshold. In this example, the value of the offer would equal 0, as on this trial, the subject would be most uncertain about the value of the offer. Thus, negatively valued offers are on trials where the offer length is more expensive than an individual's threshold (and thus are usually skipped) and positively valued offers are on trials where the offer length is less expensive than an individual's threshold (and thus are usually entered and earned). Both negatively and positively valued offers have less uncertainty as they are more easily identified as bad deals vs. good deals.

We previously characterized that the majority of trials that are quit in the wait zone comprise trials where the value of the offer was negative, where quits become the correct action to take (and subjects generally quit efficiently before too much time was invested), undoing the previous enter decision in the offer zone that should have been skipped in the first place. It remains unclear what the economic characterization is of

Figure 4.18: Modeling sub-optimality and economic efficiency across species



Proportion of total rewards mice (A), rats (B), and humans (C) actually earned relative to model-estimated maximal predicted earnings taking into account individual differences in behavioral performance and subjective valuation preferences (see Methods and Supplemental Discussion). Observations are normalized to perfectly optimal earnings in our model (1.0, e.g., using minimal reaction times and no quits based on each subject's behavior and restaurant-specific thresholds). Mice, rats, and humans all behaved significantly sub-optimally according to this model (t-test comparing means against 1.0, mice: $t=-31.962$, $p<0.0001$; rats: $t=-8.43$, $p<0.0001$; humans: $t=-20.29$, $p<0.0001$). (open circles in (A-C) represent individual animals, error bars ±1 SEM). (D-F) Pooled data across all subjects to visualize economic efficiency of quit decisions. This was measured by calculating two "value" metrics (offer value defined as the difference between each subject's restaurant-specific normalized threshold and offer, value left defined as the difference between threshold and the remaining countdown at the moment of quitting) (see Methods and Supplemental Discussion). Dashed blue lines indicate 0 value for both metrics, where either the offer = threshold (vertical) or the time remaining in the countdown when quitting = threshold (horizontal). In (F), a single example trial is circled in green and described in detail for demonstration purposes. Pie charts in (D-F) quantify percentage of all quits that fall into each of the 3 color-coded quadrants of the scatter plot (green = smart quits; red/orange = bad quits).

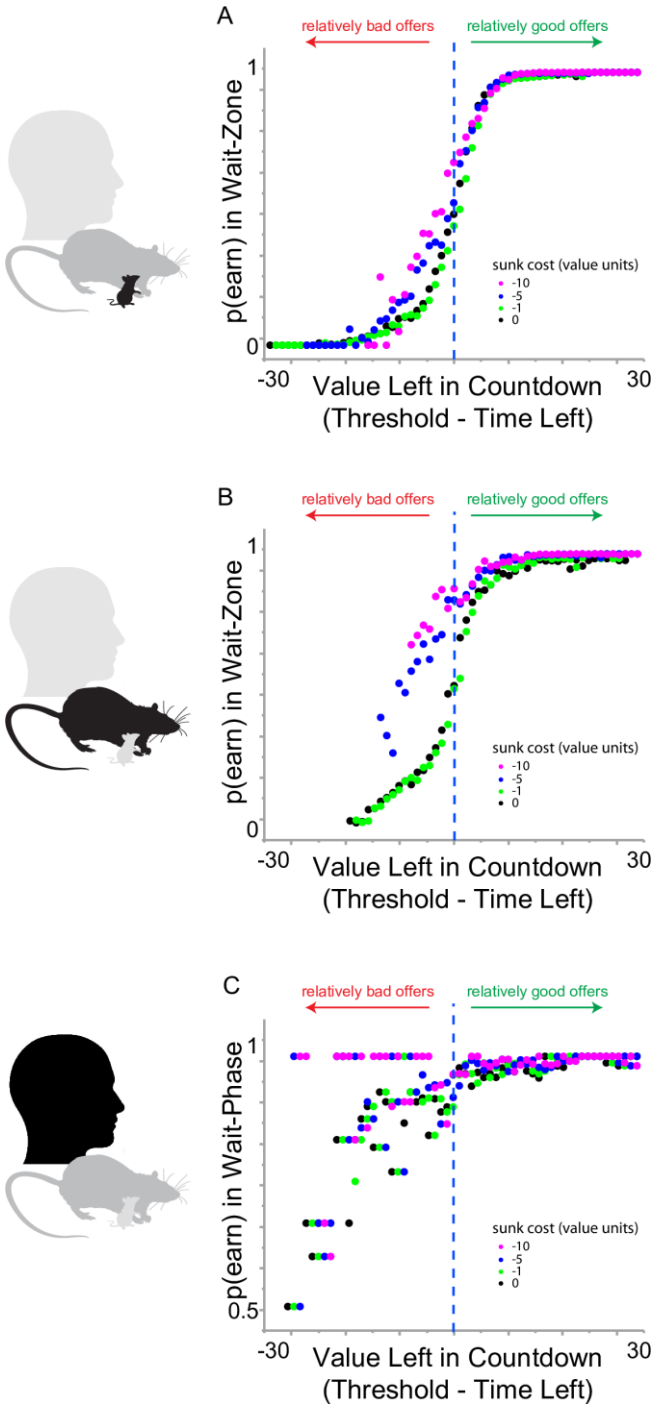
the trials where sunk costs inflate the value of rewards in pursuit and escalate level of commitment in the wait zone.

Therefore, in another set of analyses, we directly test wait zone sunk cost effects in value terms, characterizing the value of the offer accepted as well as the value remaining in the countdown at the moment of quitting in order to calculate the effects of time (in value units) invested on p(earn) in the wait zone (Figure 4.19). We found that the majority of sunk-cost driven inflation of reward value actually took place for negatively valued amounts of time left remaining in the countdown at the moment of quitting when p(earn) was calculated and when prior amounts of time were invested (Figure 4.19) In other words, the effects of time already invested only had an effect on increasing p(earn) in the wait zone if added investment history accrued while both the offer was above threshold and the time left remaining in countdown was also above threshold.

This reveals an interesting feature of the sunk cost effect that aligns with notions brought up in chapter one. Change-of-mind decisions are capable of evoking regret in subsequent trials after quit decisions were made even though those quit decisions were the economically advantageous thing to do. We previously demonstrated that mice are capable of learning from regret in order to avoid future potentially regret-inducing scenarios. That is, mice learned to actually skip those bad offers in the first place and avoid getting "stuck" in the wait zone after having accepted a bad offer only to be left torn between doing the correct thing (quitting) despite carrying the burden of regret later vs. doing the economically disadvantageous thing (stay and earn) thus committing the sunk cost error. The latter option, consistent with human reports of the sunk cost effect, too may be a way to avoid cognitive dissonance like regret.

In this analysis, because sunk cost effects only inflate the value of continued reward pursuit only during trials where offer zone mistakes were made and negatively valued offers were in question, this suggests that the sunk cost effect was present in the wait zone particularly for potentially regret-inducing scenarios had those trials been quit. Whether or not subjects committed the sunk cost fallacy to actively / intentionally

Figure 4.19: Characterizing the sub-optimality of sunk costs directly as a function of subjective value



Sunk cost analyses re-run in value terms in mice (A), rats (B), and humans (C). Black data points represent trials where the value of the offer is indicated by the x-value with no time already invested prior to when p(earn) was calculated. Vertical dashed blue line indicates zero value, where value left is at threshold. Sunk cost inflation of p(earn) is detected for negatively value left when additional time was invested prior to that point (additionally negative sunk value).

avoid regret is unclear, but this analysis reveals that sunk costs due accrue for these particular types of economic scenarios. We do know that earned rewards that were more expensive are more highly valued across species.

With regard to value uncertainty, decisions for offers near threshold (near an indifference point) are most uncertain. Thus, negatively valued offers are (usually) more readily skipped as those offers are more certainly identified as bad offers. Likewise, positively valued offers are more readily entered as those offers are more certainly identified as good offers. Interestingly, if a bad offer is accepted and subjects enter the wait zone, the way in which value-based uncertainty changes as a function of investment history opens an interesting interpretation of sunk cost effects. That is, for every additional second invested in a bad offer, the value remaining in the countdown (which started above threshold) would therefore approach threshold more and more with each passing second. What this implies is that the certainty that a bad offer is in fact bad becomes less certain with every passing second in the wait zone. This in theory could promote staying in the wait zone compared to the previous second where the value of this bad deal was more certain that one ought to quit. This change in value-based uncertainty could promote sunk cost-driven staying in the wait zone. Furthermore, if time spent waiting in the wait zone for a bad deal advances the time remaining in the countdown below an individuals threshold (i.e., maximal uncertainty), then from this point onward while waiting, the value of the reward offer would therefore become more certain that it is a good deal with each passing second. Thus, the value of staying would similarly increase as a function of investment history. Interestingly, we only observe an increase in the value of staying as a function of investment history only when the value remaining is negative and not when the value remaining is positive, again suggesting an intimate link between regret-related processes and not simply changes in value certainty.

Discussion

A sensitivity to sunk costs defies optimality considerations (Figure 4.18). So, why has this cognitive bias persisted across evolution? Three plausible psychological mechanisms that support sunk cost biases include (1) that it may be more advantageous to calculate reward value through effort expended, (2) State-Dependent

Valuation Learning (SDVL), and (3) Within-Trial Contrast (WTC) processes (Aw et al. 2011; Pompilio et al. 2006; Kacelnik and Marsh 2002; Church et al. 1991; Wikenheiser et al. 2013; Pattison et al. 2012). We discuss each of these below.

Because predicting valuations that depend on future outcomes is complex and difficult, animals may have evolved processes in which valuation is measured from effort spent rather than calculated as an estimate of unknown future outcomes. Past effort is easy to measure but has a limited (but non-zero) correlation with future value. In contrast, calculating value from expected future outcomes has its own estimation uncertainties. If the correlation between past efforts and future value is higher than the uncertainties of future outcomes, then animals may have evolved processes that use past effort as a proxy to estimate future value (Pompilio et al. 2006; Wikenheiser et al. 2013). This can explain our observation that the post-consumption evaluation increases proportionally to the time spent waiting for the reward in all three species (Figure 4.4). The fact that susceptibility to sunk costs only accrued in the wait zone implies that valuations in the offer zone depend on different processes that do not include measures of effort spent, but which may be more related to direct estimates of future value.

The State-Dependent Valuation Learning (SDVL) theory hypothesizes that energy spent working toward reward receipt moves the individual into a poorer energy state, enhancing the perceived value of the yet-obtained reward (Pompilio et al. 2006; Aw et al. 2011). This continued work can thus escalate commitment of continued reward pursuit with growing sunk costs. Similarly, the Within-Trial Contrast (WTC) theory describes the sunk cost phenomenon as increasing contrast between the decision-maker's current physical state and the goal (Singer and Zentall 2011). SDVL and WTC propose that either physiological or psychological states could drive added value leading to a susceptibility to sunk costs. However, we did not observe sunk costs accruing during the offer zone, even though time spent in the offer zone is equivalent in physical and cognitive demands to time spent in the wait zone. Simple explanations from WTC and SDVL would predict sunk costs to accrue in the offer zone as well.

Past-effort heuristics, SDVL, and WTC can indeed be prominent drivers of the sunk cost effect in our data when sunk cost effects are present. Therefore, our work brings up an intriguing question: what is different in decision-making processes between those made in the wait zone (susceptible to sunk costs) and those made in the offer zone (not susceptible to sunk costs)?

One possibility is that decisions made in the offer zone and wait zone may rely on separate processes that calculate value in distinct ways through dissociable neural circuits (Redish 2013; Meer et al. 2012; Casey et al. 2008). Recent findings from other foraging tasks suggest that choosing to remain committed to already accepted options accesses different valuation algorithms than deliberating between distant options (Wikenheiser et al. 2013; Redish et al. 2016; Carter and Redish 2016; Stephens and Krebs 1986). We suggest that offer zone decisions are driven by forward-looking deliberation mechanisms, while wait zone decisions are driven by distinct mechanisms that depend on information about accumulated past states. There is strong neural evidence across species to suggest that competing future alternatives are being represented, evaluated, and compared in deliberation algorithms (including during VTE behaviors in rodents, Figure 4.12), while other decision-making systems depend on past-self representations (Meer et al. 2012; Redish 2013; Wikenheiser et al. 2013; Papale et al. 2016; Bushong et al. 2010; Dayan et al. 2006). Each of these decision-making systems provide computational advantages better suited for different situations. Thus, these multiple valuation algorithms can each confer independent evolutionary advantages and can co-exist and persist across time and species (Casey et al. 2008; Meer et al. 2012; Redish 2013).

The sunk cost fallacy, by definition, arises from valuing spent resources that cannot be recovered. Our data finds that these sunk costs only accrue under specific situations in mice, rats, and humans. Past studies reporting conflicting findings across species may have failed to consider the nature with which different decision-systems drive behavior. We suggest that multiple parallel decision-making valuation algorithms implemented in dissociable neural circuits have persisted across species and over time through evolution. Our data imply these different valuation algorithms are differentially susceptible to sunk costs.

Using a translational approach in mice, rats, and humans, we find direct evidence in parallel tasks that the sunk cost phenomenon is conserved across species. Our findings highlight the utility of economic paradigms that can dissociate decision-making computations using naturalistic tasks that are translatable across various species and can be expanded to survey individuals of varying ages or psychiatric populations.

There is a large body of literature sparked by an early publication from Arkes & Ayton that not only suggested non-human species are incapable of displaying the sunk-cost fallacy, but that human children are also incapable of displaying the sunk cost fallacy (Arkes and Ayton 1999). That is, children "outperform" adults by not factoring in added value due to sunk costs. This study postulated that children have not yet learned a socionormative "do not waste" rule in these early stages of development. The majority of studies building off of this work have primarily focused on phenomenology of the sunk cost fallacy in the context of aging and have largely focused on developmental processes from a psychology perspective, departing from comparative studies across species as well as advancements in neuroscience. As a result, the original notion put forth by Arkes & Ayton has remained largely unchallenged, with caveats, exceptions, and inconsistencies found in follow-up human studies informing newer theories of developmental psychology that further depart from cross-species and neurobiological mechanistic work on decision-making information processing (Arkes and Ayton 1999).

For instance, work on impulsivity in children is mixed and at odds with Arkes & Ayton's claim that children are more likely to resist sunk costs (Arkes and Ayton 1999). Several studies in children report sometimes positive, sometimes negative, or sometimes no correlation between decision-making competency or impulsivity inventory scales and susceptibility to honor sunk costs (Bruin et al. 2007; Parker and Fischhoff 2005; Weller et al. 2012). Various explanations attempting to explain such discrepancies have been proposed, rooted in self-regulation, perseveration, learned rules to not waste, and ability to recognize when to activate appropriate heuristics. (Weller et al. 2012; Dhami et al. 2011; Klaczynski 2009; 2004; 2001b; 2001a; Morsanyi and Handley 2008; Reyna and Ellis 1994; Reyna and Farley 2006). Such decision-making competency and impulsivity inventories have been well-validated, control for general mental ability and

intelligence, and are capable of predicting, for example real-world negative socioeconomic and health outcomes attributed to lifestyle choices, including risk of substance abuse and delinquency later in life (Giancola and Tarter 1999; Giancola and Mezzich 2003). Other studies show, at the other end of the spectrum, that older adults too are less sensitive to sunk costs than younger adults, adding an inverted-U shape trajectory to the variable prevalence of this cognitive bias (Bruin et al. 2014). Such reports suggest that older adults are less likely to ruminate on past expenditures and prior negative events, are better at coping with failed plans, and are less likely to mention sunk costs when making decisions about the future (Bruin et al. 2012; 2007; Strough et al. 2016; Bruin et al. 2014; Strough et al. 2008; 2011a; 2011b; 2014; Torges et al. 2008). None of these studies include any mechanistic data nor interpretation of their work in the context of changing neural circuits of different, parallel decision-making valuation systems. We propose that future studies should integrate biologically plausible models of the neural mechanisms underlying nonlinear developmental changes in separable decision systems (Redish 2013; Meer et al. 2012; Casey et al. 2008). Our data can offer a fresh lens through which we and others can revisit past literature and inform future experiments.

Taken together, these tasks and findings may shed light on novel diagnostic or intervention strategies and reveal the roles of neurally-distinct decision systems for future research in education or neuropsychiatry.

– Interlogue –

Chapter 5

Distinct valuation algorithms conserved across evolution

In the previous chapter, I demonstrated how, even through behavior, the types of information that are represented in dissociable decision-making modalities can be revealed. In this particular case, building off of the discoveries in chapter one, I explored how hidden costs and hidden utility in different decision-making systems are processed similarly in mice, rats, and humans. Separate behaviors in the offer zone and wait zone reveal fundamentally distinct valuation processes conserved across species.

As characterized in chapter one, high-conflict scenarios in this novel variant of the Restaurant Row task – encountering an expensive offer for a preferred flavor – are capable of capturing interesting episodes of competing valuations between "wanting" on the one hand vs. "knowing better" to explore elsewhere on the other hand. Importantly, I reveal how this conflict is processed is fundamentally different in the offer zone and wait zone. In chapters one and two, I demonstrate that economic mistakes in the offer zone, or giving in to waiting and entering the wait zone leave mice, rats, and humans with an interesting neuroeconomic dilemma once in the wait zone.

Once an expensive offer is accepted and a mistake has been made, subjects are left with recurring secondary re-evaluative process in the wait zone during the reward countdown where with every passing second the decision to stay vs. opt-out and abandon on-going investments is being re-considered. In chapter one, I characterized that changing one's mind in the wait zone is the economically advantageous decision to make.

By changing one's mind, one realizes that such an offer should have been skipped in the first place through counterfactual processing. Thus, mice were left with regret-like experiences moving forward.

In chapter two, I demonstrated that change-of-mind decisions in the wait zone were uniquely susceptible to the effects of accruing sunk costs such that mice, rats, and humans all inflate the value of staying in the wait zone as a function of investment history. That is, after accepting an expensive offer in the wait zone, quit-induced regret vs. sunk-cost-driven perseveration appeared to be intimately related. After making a principal judgement in the offer zone to accept an offer at a cost higher than subjective value indicates one should, subjects were faced with a change-of-mind dilemma torn between irrationally waiting out the expensive offer vs. rationally back-tracking and changing their plans, where affective contributions appear to weigh these options against one another.

Interestingly, I discovered that investment history made while deliberating about whether or not to accept an offer in the first place, even if depleting limited resources and even if contributing toward reward receipt, were not accrued as sunk costs and did not inflate reward valuations to say while waiting. This suggested that the valuation processes used while initially deliberating, or choosing-between options, factored in different information that did not lead to the escalation of commitment unlike foraging valuations to opt-out of already committed decisions.

Prominent theories in evolutionary biology that have attempted to explain why human and non-human animals alike show evidence of the sunk cost fallacy have fully considered the decision-making systems that could give rise to such cognitive biases under the neural constraints of a biologically plausible model. Information processing that is susceptible to sunk costs must factor in past and current representations of self-states. Given our findings that reveal certain decision modalities are insensitive to sunk cost driven valuations (i.e., offer zone deliberative algorithms), I suggest that the information represented in those algorithms take into account different pieces of stored information that is fundamentally distinct from decision modalities that are sensitive to sunk cost driven valuations (e.g., wait zone foraging algorithms). I

rest these interpretations on numerous behavioral, electrophysiology, and imaging studies reported across species that reveal information about future states, and not necessarily past states, are represented during forward-looking deliberative valuation algorithms.

Taken together, our discoveries suggest that the neural systems driving deliberative valuations are fundamentally distinct from those driving foraging valuations, particularly when engaged in high-conflict scenarios related to self-control. In the novel variant of the Restaurant Row task, this is reflected in dissociable behaviors separably measurable across space and time in the offer zone and wait zone.

In the next chapter, I will discuss a new theoretical framework that takes what we can learn from neuroeconomics in order to gain a fresh perspective on how different decision-making processes might be susceptible to distinct circuit-computation-specific malfunctions or failure modes underlying heterogeneous neuropsychiatric disorders.

Chapter 6

From memory to decision making: Addiction as a heterogeneous disease of computation-specific valuation algorithms

Abstract

Addiction is considered to be a neurobiological disorder of learning and memory because addiction is capable of producing lasting changes in the brain. Recovering addicts chronically struggle with making poor decisions that ultimately lead to relapse, suggesting a view of addiction also as a neurobiological disorder of decision-making information processing. How the brain makes decisions depends on how decision-making processes access information stored as memories in the brain. Advancements in circuit-dissection tools and recent theories in neuroeconomics suggest that neurally dissociable valuation processes access distinct memories differently, and thus are uniquely susceptible as the brain changes during addiction. If addiction is to be considered a neurobiological disorder of memory, and thus decision-making, the heterogeneity with which information is both stored and processed must be taken into account in addiction studies. Addiction etiology can vary widely from person to person. I propose that addiction is not a single disease, nor simply a disorder of learning and memory, but rather a collection of symptoms of heterogeneous neurobiological diseases of distinct circuit-computation-specific decision-making processes.

Chapter reprinted with permissions from *Learning & Memory*, modified from:

Sweis BM, Thomas MJ, Redish AD. 2018a. Beyond simple tests of value: Measuring addiction as a heterogeneous disease of computation-specific valuation processes. *Learning & Memory* (in press).

Introduction

The National Institute on Drug Abuse (Volkow 2012) defines addiction as "…a chronic, relapsing brain disease that is characterized by compulsive drug seeking and use despite harmful consequences. It is considered a brain disease because drugs change the brain – they change its structure and how it works. These brain changes can be long-lasting and can lead to the harmful behaviors seen in people who abuse drugs." Thus, addiction has been proposed to be a neurobiological disorder of learning and memory because drugs of abuse can leave lasting changes on the structure and function of the brain (Hyman 2005). These changes are thought to underlie why individuals with addiction struggle with making poor decisions.

Building off of the concepts of multiple memory systems and multiple decision-making systems I previously discussed in Chapter 1 – if addiction is to be considered a neurobiological disorder of memory, it must also be considered a neurobiological disorder of decision making and the heterogeneity with which information is both stored and processed must be taken into account. Here, I review the complexity and diversity of addiction at the intersection of mechanisms of circuit-specific memory and multiple decision-making information processing systems. Given the recent advancement of circuit dissection tools, our approach makes predictions as well as suggestions for the field moving forward in efforts toward developing novel computation-specific therapeutic interventions.

On addiction heterogeneity

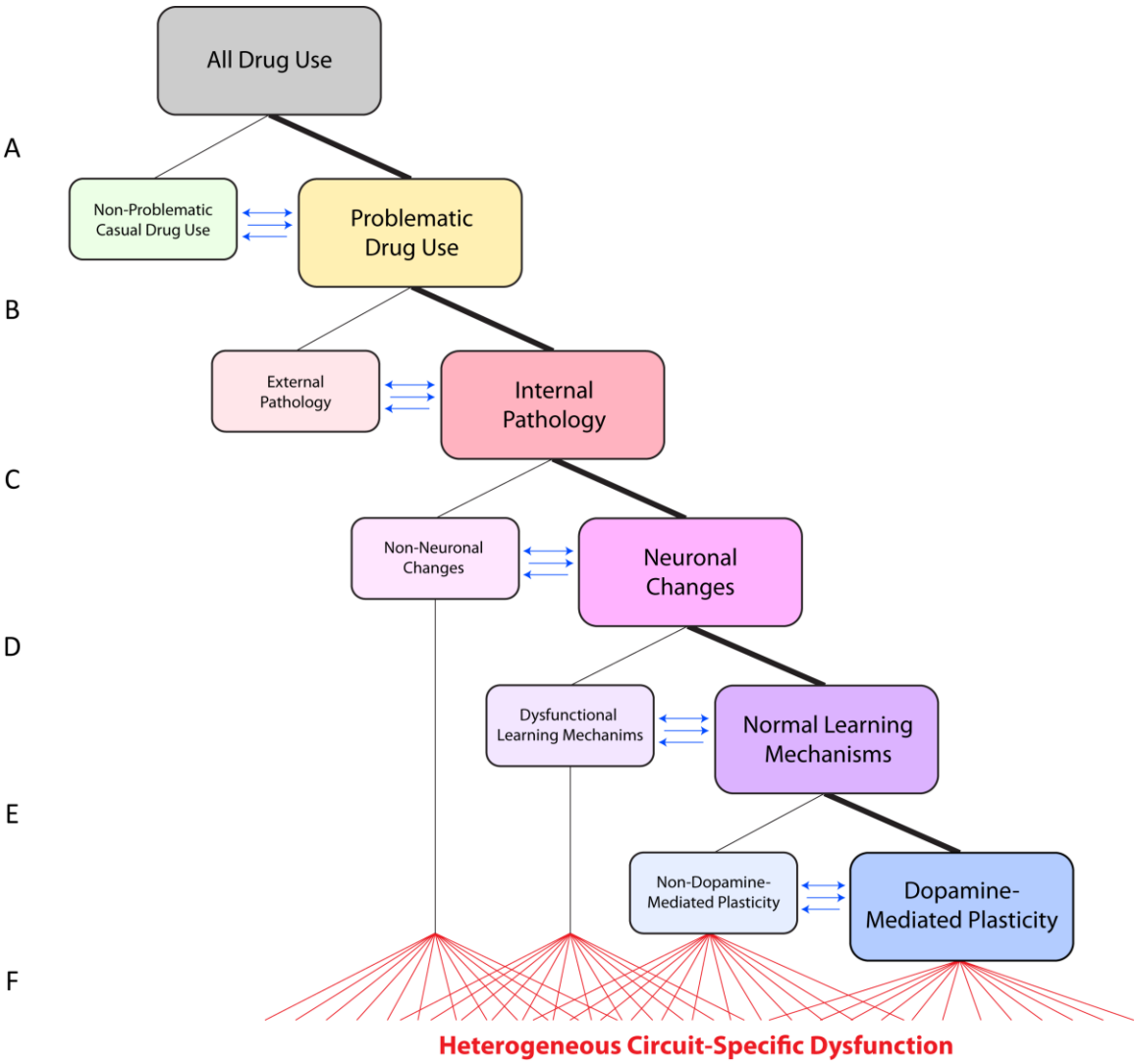
Drug-related experiences are capable of leaving lasting influences on the brain and behavior (Le Moal and Koob 2007; Robinson and Berridge 2003). Thus, problematic drug abuse and susceptibility to relapse is often attributed to neuronal pathologies in mechanisms of learning, memory, and plasticity (Nestler 2001; Hyman and Malenka 2001; Hyman 2005; Kauer and Malenka 2007; Lüscher and Malenka 2011; Thomas et al. 2008). This view has dominated much of addiction neurobiology research. However, an ultimate understanding of addiction pathogenesis will depend on appreciating the multiple plausible causes of drug-use (Figure 6.1).

The vast majority of drug users in fact do not go on to display problematic drug use (Figure 6.1A) (Anthony et al. 1994). Comparisons between casual drug users and problematic drug users can perhaps reveal fundamental differences in neurobiological functions underlying the severity and chronicity of or resiliency from relapse. There is a large body of work in both human and non-human animals comparing compulsive drug-seeking behaviors to non-drug-exposed and non-drug-treated controls. There is a relatively smaller literature comparing individual differences in compulsive drug-seeking behaviors to casual drug consumption.

However, several recent studies have suggested that even within standard animal addiction models, only subsets of animal subjects show phenotypes that closely resemble human addictions (Ahmed 2005; 2010; Ahmed et al. 2013; Pickard et al. 2015). For example, Vandaele et al (2016) found that although all rats tested showed larger progressive ratio breakpoints to cocaine and heroin than to saccharin, most rats would choose saccharin over cocaine or heroin if given an actual choice between them. Perry et al (2013) found that the subset of rats that would choose cocaine or heroin over saccharin in the two-available-choice condition showed other similar addiction-related phenotypes. Deroche-Gamonet et al (2004; see also Kasanetz et al. 2010) found that, while all rats would lever-press or nose-poke for cocaine, only a subset would cross a shock to reach the drug, and that same subset showed excessively high progressive ratio breakpoints. Jaffe et al (2014) found that the subset of nicotine-seeking rats that showed excessively high progressive ratio breakpoints also showed a lack of Kamin blocking response to nicotine, even though their Kamin blocking response to food was normal.

Most studies investigating the neuroscience of addiction pathology have focused almost exclusively on those due to internal factors within a subject, meaning within the brain (Figure 6.1B). Pathology, by definition, is derived from the Greek words “pathos” meaning “disease” and “logos” meaning “treatise,” describing the science of the causes and origins of disease states. Sometimes the term pathology can be misused and instead of referring to disease origin, may sometimes be ascribed to reflect measurable

Figure 6.1: On addiction heterogeneity: Classes of plausible dysfunctions.



(A) All drug use can be sub-divided into casual drug use (majority) vs. problematic drug use. (B) Individuals with problematic drug use can be divided into those with pathologies (originating cause) due to external, social factors vs. pathologies rooted within the individual either via a predisposing vulnerability or a direct change induced by an ingested substance. (C) Internal pathologies can be divided into those with primary changes in neurons vs. non-neuronal players (e.g., glia). (D) Neuronal changes, or plasticity, can be divided into changes that come about from normal mechanisms of learning and memory or a dysfunctional breakdown of such processes that normally do not occur. (E) Normal mechanisms of learning and memory can be driven in reward-related circuits by dopamine-mediated processes or non-dopamine-mediated processes (e.g., endocannabinoid signaling). Blue arrows in between nodes indicate interaction pathologies that could have either unidirectional or bidirectional influences on each other. (F) Any resultant changes in the brain that arise from internal pathologies, regardless of the underlying primary mechanism, can each induce failure modes in dissociable neural circuits, each of which can give rise to fundamentally distinct addiction etiologies in separable neural computations.

evidence of disease state, either as dysfunctions downstream of disease origin or even disease symptomology. However, in certain cases, the cause and origin of problematic drug use can be entirely attributed to external (i.e. environmental) factors. For example, in a neural system that may be working perfectly normally, with no predisposition or known susceptibility to drugs, it might be the case that associations constructed in certain social settings can drive drug abuse. For instance, television ads or the entertainment industry that publicize, glorify, or sexualize drug use can drive problematic drug use in viewers with access to those media. In such cases, a failure in an individual's brain may not be the underlying origin of his or her addiction, nor might the mechanism of action of the ingested substance on the individual's brain be the origin of maladaptive behavior. In such cases, successful treatments may need to be rooted in environmental and social interventions. Of course, two individuals who undergo similar media exposures might yet emerge with different disease states. Thus, for such questions, problematic drug use is likely rooted in an interaction pathology between the individual neural system and the external environment (Figure 6.1B blue arrows).

Many of the neuroscience studies of addiction and the brain have focused almost exclusively on those with neurophysiological etiologies (Figure 6.1C). However, recent work highlights the involvement of key non-neuronal players including glia and the immune system, both centrally and peripherally located, that could serve roles as sources of dysfunction, potential diagnostic or prognostic disease biomarkers, or possible novel avenues for therapeutic intervention. Astrocytes and microglia functions, including those critical for synapse formation, can be disrupted by alcohol, psychostimulants, and opioids through either direct means, indirect means through neuronal signaling, or through the innate immune system, which together contribute to added drug-abuse liability (Navarrete and Araque 2008; Lacagnina et al. 2017; Calipari et al. 2018; Lewitus et al. 2016; Northcutt et al. 2015; Araos et al. 2015; Karlsson et al. 2017; Fox et al. 2012). It is certainly possible that primary dysfunctions in non-neuronal players could give rise to secondary neuronal dysfunctions and vice versa with interacting pathologies (Figure 6.1C blue arrows), but it is important to consider primary dysfunctions in order to better appreciate addiction etiology. It is possible that both neurophyisological and glial dysfunctions could exist in fundamentally distinct forms of addiction, yet this remains to be explored. One could imagine a scenario in which primary dysfunction in glia that give rise to maladaptive neuronal

plasticity could continue to give rise to maladaptive circuit changes even if neuron-targeted therapies are administered.

Within the realm of neuronal changes, some addiction-related changes depend on dysfunctional forms of plasticity and some are instantiated through normal learning mechanisms (Figure 6.1D). Addiction-related changes in memory are often thought to be functioning within intact learning mechanisms at the molecular and cellular levels (if perhaps at an accelerated rate or enhanced level) (Hyman and Malenka 2001; Hyman 2005; Lüscher and Malenka 2011; Kauer and Malenka 2007; Hearing et al. 2018; Thomas et al. 2008). The vast majority of nonhuman-animal studies examining these questions start from the hypothesis that drugs of abuse, either directly or indirectly, take advantage of endogenous reward-related systems and usurp normal mechanisms of learning and memory. While a prominent view, there are several reports suggesting this is only one potential path to drug addiction. In fact, it is important to be careful with the term "plasticity." Changes in strength of synaptic transmission (e.g., potentiation, depression) either through pre-synaptic mechanisms (e.g., changes in vesicle release probability) or post-synaptic mechanisms (e.g., changes in receptor densities) can store information about the historical past (Thomas et al. 2008). Changes in plasticity itself, often termed "metaplasticity," describe a distinct process in which the history of a given structure alters the direction or magnitude of plasticity in response to subsequent stimulation (Abraham and Bear 1996). Thus, metaplasticity describes changes that augment or diminish the overall degree with which a system is capable of changing and thus can interact with normal learning mechanisms (Figure 6.1D. blue arrows).

The vast majority of neuroscience addiction research has focused on changes in dopaminergic signaling or synaptic pathways known to be modulated by dopamine. (Figure 6.1E). However, plasticity involving other neuromodulators, including endocannabinoids, serotonin and norepinephrine are known to contribute to learning and memory, and any of these might also be subject to change in addiction (Chevaleyre et al. 2006, Weinshenker and Schroeder 2007, Müller andHomberg 2015) and could conceivably interact with each other (Figure 6.1E blue arrows).

In light of these multiple plausible causes underlying continued drug use, a key concept to consider is to what extent might heterogeneous circuits be differentially affected (Figure 6.1F)? Even within a single branch of plausible neurobiological mechanisms of addiction – changes mediated through dopaminergic pathways – studies examining drug-induced plasticity have not fully appreciated the heterogeneity of circuits within which these mechanisms may be taking place, especially within the context of decision-making information processing. That is, questions of which circuits and what information underlie addiction-related plasticity remain largely unresolved.

On circuit heterogeneity

In order to probe the multiple memory systems and multiple decision-making systems in the brain, circuit-specific and temporally precise tools that can gain access to endogenous mechanism of both information storage and information processing are required. Recent advances in circuit-dissection tools have made studying specific populations of cells possible (Yizhar and Adamantidis 2018). Conditional genetics have allowed experimenters to define interrogation parameters based on cell-expression profiles, activity-dependence, and projection specificity, leveraging voltage- and calcium-sensing reporters with novel imaging or opto-tagging techniques and utilizing chemogenetic or optogenetic manipulations (Resendez et al. 2016; Guo et al. 2015; Kim et al. 2017). As a result, newer studies have been able to increase the functional resolution of diverse circuit-specific specialization.

Recent delineation of circuit heterogeneity of the mesocorticolimbic dopaminergic system illustrates this point. Novel circuit-dissection tools have revealed the wide diversity of unique inputs into the VTA on dopaminergic and non-dopaminergic neurons as well as the wide diversity of unique VTA output structures that include the cell-type identity of their targets (e.g., medial VTA dopaminergic neurons that receive glutamatergic input from the lateral habenula and project to D1-receptor-containing GABAergic medium spiny neurons in the shell of the nucleus accumbens) (Morales and Margolis 2017). Pathways thought to be critically affected by addiction are being sub-divided at a rate faster than the functional roles of those different sub-circuits are being realized (Britt et al. 2012; Tye 2012; Morales and Margolis 2017).

With these tools available, we suggest that a critical factor to aid in our understanding of the neurobiology of addiction as the field moves forward requires taking into consideration the multiple memory and multiple decision-making systems. Importantly, this means not only moving beyond simple tests of value, which are commonplace in many addiction studies, but also designing the circuit-manipulation approach so as to carefully probe the link between memory and decision-making processes. To accomplish this, we argue that direct "off-line" manipulations of circuit-specific plasticity can help realize the functional consequences of synaptic remodeling on endogenous circuit-computation-specific information processing. Because lasting heterogenous aspects of memory can be altered due to a variety of addiction-related pathologies, such changes can give rise to unique susceptibilities to addiction via vulnerabilities in circuit-specific computational processes that might otherwise be masked either by using simple tests of value or by using circuit interrogation techniques that disrupt endogenous information processing. We discuss these concepts further below.

Moving beyond simple tests of value

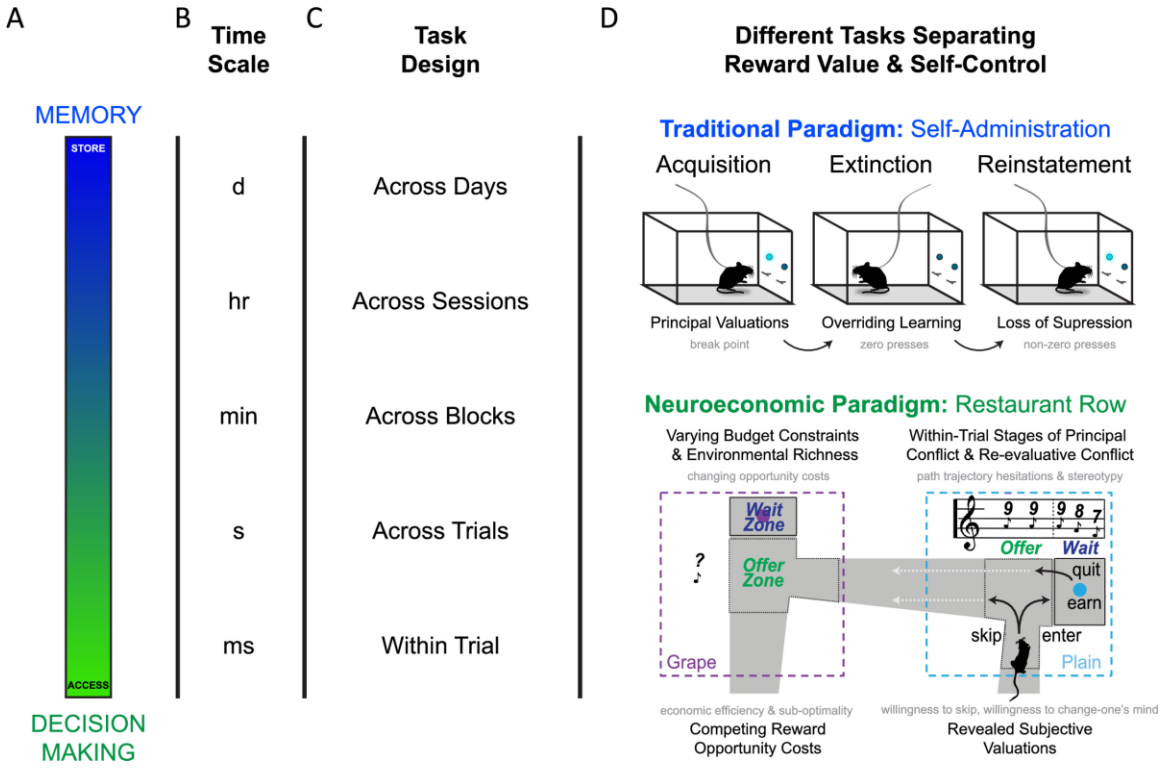
Over the last four decades, behavioral neuroscience has developed a variety of tasks that can separate these decision-systems, by putting them into conflict with each other. For example, classic studies of devaluation that find differences between different training contingencies (Balleine and Dickenson 1998; Coutureau and Killcross 2003), or studies of water maze behavior that depend on training (Morris et al 1982; Eichenbaum et al 1990; Day and Langston 2006; Redish 1999), or the classic T-maze and other contingency-dependent tasks (Barnes et al 1980; Packard and McGaugh 1992; Schmidt et al 2013; Gardiner et al 2013). In more recent work, neuroeconomics has refined these tasks and developed additional tasks that can measure the valuation of these components directly (Coricelli et al. 2005; McCoy and Platt 2005; Hayden et al. 2008; 2009; Abe and Lee 2011; Kalenscher and van Wingerden 2011; Steiner et al 2014; Sweis et al 2018b, 2018c; Chapters 2 & 4).

In contrast, many of the tasks used in rodent models of compulsive drug-seeking behaviors have relied on paradigms that, by design, are better suited to probe the changes in information stored (i.e. mechanisms of memory), rather than the changes in behavioral processing (i.e. decision-making) (Figure 6.2). Such tests

include drug conditioned place preference or drug self-administration paradigms (Tzschentke 2007; Stafford et al. 1998). In these paradigms, time voluntarily spent in a drug-paired chamber, or, number of drug infusions self-administered on a progressive ratio schedule serve as the primary behavioral metrics of reward valuation in these tasks. However, additional valuation information can be extracted from these tasks, examining rate of learning acquisition, individual differences in high-responders vs. low-responders, or measuring how such behaviors can be extinguished and reinstated following a number of various triggers (Piazza et al 1990; Carroll and Lac 1993; Gosnell 2000; Lu et al. 2003; Perry et al. 2005; Shaham et al 2003). These latter examples have been used to model "relapse" in non-human animals and have provided a foundational understanding about the neurobiological mechanisms of reward-related learning and memory associated with addiction.

When addiction models have been tested with tasks derived from behavioral neuroscience, conflicting results have often been found. For example, rats that were willing to expend more lever presses in a progressive ratio schedule for cocaine or heroin than for saccharine, still chose the saccharine when given an option between the two (Ahmed 2005; 2010; Ahmed et al 2013; Perry et al. 2013; Vandaele et al. 2016), consistent with experiments in monkeys that found differences between separated, bundled, and contrasted options (Nader and Woolverton 1991; 1992; Nader et al. 1993; Czoty et al. 2005; John et al 2015). While most rats would take cocaine if it was the only option available, only a subset showed a willingness to cross a shock for it (Shalev et al. 2000; Deroche-Gamonet et al. 2004; Belin et al. 2009; Deroche-Gamonet and Piazza 2014; Martin-Garcia et al. 2014). While many studies have shown that drugs (cocaine, alcohol, heroin) do not respond to devaluation (Dickinson et al. 2002; Everitt and Robbins 2005; Zapata et al. 2011; Everitt 2014), suggesting they are due to procedural and habit processes, other studies have shown that subjects can plan complex novel sequences to drug rewards (Olmstead et al. 2001) and that rats will use knowledge of historical valuations to plan options (Marks et al. 2010).

Figure 6.2: Tasks design matters when probing memory vs. decision-making processes



(A) Memory and decision-making are thought to exist as duals of each other. How information is stored changes how it is processed. Different decision-making mechanisms access stored information traded off in different ways, and thus, select actions by fundamentally distinct computational algorithms. (B) Tasks that interrogate processes on varying time scales are better suited to probe memory vs. decision-making computations. (C) Tasks designed measuring behaviors on those longer time scales (days) versus shorter time scales (within trial) are better suited to probe memory mechanisms (information storage, consolidation, updating) vs. decision-making mechanisms (information processing and action-selection valuations). (D) Two task examples better suited to probe either memory processes (traditional paradigm: self-administration task) and decision-making processes (neuroeconomic paradigm: restaurant row task), both of which are capable of investigating aspects of reward-related self-control. Top: In traditional operant chamber paradigms, principal or initial reward valuations (for food or drug) are measured during an acquisition learning period (usually across trials or days, estimated via break point on a progressive ratio lever-press sequence). Extinction periods can probe rates with which new valuation processes are learned that suppress principal valuations (across days). Active maintenance of extinction learning, or susceptibility to lose suppression following reinstatement, implies principal valuation memories co-exist with extinction memories yet such competing computations are not accessible in traditional operant paradigms. Bottom: In neuroeconomic paradigms, reward value can be calculated a number of ways within a single trial. In this version of the restaurant row task, hungry mice are trained to forage for food rewards of varying costs (delay, cued tone pitch) and subjective value (flavor, spatial contexts or restaurants) while on a limited time budget. Decisions are deconstructed into discrete stages in separate offer zones and wait zones on each trial in each restaurant. Each action-selection process reflects a valuation computation, each of which reflect different economic algorithms (choose between entering vs. skipping in the offer zone, deciding to opt out and quit vs. remain patient until earning in the wait zone, taking time to consume a pellet and linger in a conditioned place vs. leave and advance to the next trial). In each of these action-selection processes, decision conflict and self-control can be separately captured between highly desired although expensive reward opportunities.

Recent theories in neuroeconomics suggest that decisions made in different situations derive from different valuation functions residing in separable neural circuits. It can be difficult to segregate these parallel information processing algorithms using traditional experimental addiction paradigms that rely on simple tests of value and compulsive drug-seeking behavior. Even within the same trial, decision algorithms can change and thus the computations ultimately driving behavior can be multi-faceted (Sweis et al 2018b, 2018c; Chapters 2 & 4). For instance, the value assigned to choosing one option over another can be calculated through distinct discounting functions depending on what information is being incorporated in a given decision (e.g., intertemporal deliberative choices deciding between "this option or that option" vs. foraging choices deciding "to give up on and abandon the current endeavor", Carter and Redish, 2016, Sweis et al. 2018b, 2018c). In this light, distinct neural dysfunctions in either of these processes could ultimately lead to maladaptive behavioral consequences that might be indistinguishable in simple experimental paradigms.

Reinforcement learning protocols, in the form of drug conditioned place preference and self-administration, provide useful information about mechanisms of drug memory consolidation and retrieval by measuring acquisition, extinction, and reinstatement parameters. These processes are revealed across sessions on the timescale of days to weeks. Furthermore, these tasks can characterize how existing memories change slowly over time when contingencies are updated. This is clearly demonstrated when reward-seeking behaviors are extinguished over time (across days). This is also seen in reinstatement sessions when learning is updated and drug-seeking behaviors re-emerge in models of relapse.

For example, numerous studies of extinction-reinstatement processes have determined how extinction learning processes acquire new valuations (in this example, "not to seek drug rewards") that override existing, originally learned drug-seeking valuations. Importantly, it is widely accepted that this form of learning is acquired in addition to existing valuations rather than a process in which old valuation learning is removed or forgotten (Bouton 2004). While these separable memory systems are indeed thought to occur in distinct neural circuits (Berman and Dudai 2001; Suzuki et al. 2004; Peters et al. 2009), it is unclear how distinct decision-making systems gain accesses to such separate memories during on-going action-selection

processes. In the remainder of this thesis, I will present new discoveries that demonstrate how a taking a neuroeconomics approach and moving beyond simple tests of value can reveal hidden circuit-specific failure modes in decision-making information processing in animal models of addiction that would otherwise be masked by similar-appearing overt changes in behavioral output (Chapters 7 and 9). Below, I will describe how thinking about multiple decision systems in the context of multiple memory systems can inspire new ways to think about probing the neural circuits involved in healthy and maladaptive behavioral processes.

In simple behavioral paradigms, for instance, following the extinction of conditioned place preference, it is accepted that originally learned valuations are not simply removed or forgotten but rather new overriding valuations are secondarily learned and that both forms of memory remain intact (Rescorla 2001; Bouton 2004). However, it remains unclear how both types of stored information compete with one another since they are thought to co-exist. That is, how is competition between multiple decision-making systems resolved before an action is ultimately selected or not? Regardless if extinction is maintained or if reinstatement is precipitated, how are these separately stored reward-related memories accessed and integrated to produce a single behavioral output? The decision to leave the reward-paired chamber vs. not to leave could reflect fundamentally distinct neural computations from a decision to remain in the unpaired chamber vs. actively seeking out the reward-paired chamber (German and Fields 2007b; 2007a). These separate computations could be differentially disrupted in distinct forms of addiction. A similar argument could be made among various subtle action-selection processes in drug self-administration tasks (e.g., differences in trained lever presses vs naturalistic nose pokes) (Gerhardt and Liebman 1981).

Cross-task comparisons imply that hypotheses of singular, objective definitions of reward value are problematic and lead to economic paradoxes. For instance, reward value measured by calculating the breakpoint of lever pressing in a progressive ratio self-administration paradigm for drug in one session vs. saccharin in a separate session is inconsistent with reward value measured in the same animals by calculating the probability of choosing drug over saccharin in a two-alternative forced-choice paradigm (Vandaele et al. 2016; Perry et al. 2013; Deroche-Gamonet et al. 2004; Kasanetz et al. 2010). Choosing between options is

thought to access fundamentally distinct processes from choosing to remain committed to vs. abandoning current endeavors (Carter and Redish 2016; Redish et al. 2016). The former is thought to recruit deliberative processes (Redish 2016) while the latter is likely driven by Pavlovian associations embedded in foraging processes (Dayan et al. 2006). Thus, reward value is not singular within the brain, but rather depends on the separate neural algorithms used to compute value in distinct decision systems. By moving beyond simple tests of value, complex tasks may be better able to operationalize reward value in a number of various ways, sometimes within the same task, and even within the same trial (Figure 6.2). This allows neural computations to be more readily dissociable during on-going decisions.

Tasks rooted in theories of fundamentally distinct valuation algorithms are capable of dissociating neural computations underlying specific aspects of decision-making information processing, including multiple dimensions of reward value (e.g., effort, price, opportunity cost, reward magnitude, budget constraints [Wikenheiser et al. 2013; van Wingerden et al. 2015; Salamone et al. 2018; Sweis et al. 2018b; 2018c; Chapters 2 & 4]) separate from other behavioral processes (e.g., locomotor capabilities, spatial and semantic knowledge of task rules and contingencies), accessing revealed subjective valuations rather than assuming reward magnitude objectively (e.g., using reward quality and individual preferences, not reward quantity [Levy and Glimcher 2011; Steiner and Redish 2014, Sweis et al. 2018b; 2018c; Chapters 2 & 4]), introducing changes in external demands (e.g., leaner vs. richer environments [Wikenheiser et al. 2013; Sweis et al. 2018b; Chapter 2]), and deconstructing stages of decision-making discretely within trial (e.g., stepwise information presentation and action-selection processes leading up to reward earning separated across space and time [Sweis et al. 2018b; 2018c; Chapters 2 & 4]). These novel task designs can access economic principles in human decision-making (both those well-studied and those still perplexing; e.g., theories of demand elasticity and compensation, counterfactual processing, regret, sunk costs, post-purchase rationalization, and other cognitive heuristics [Camille et al 2004; Coricelli et al. 2005; Hayden et al. 2009; Abe and Lee 2011; Steiner and Redish 2014; Sweis et al. 2018b; 2018c; Chapters 2 & 4]). Importantly, neurally dissociable decision-systems, including Pavlovian, deliberative, and procedural processes can be

more closely tracked through behavior separated across space and time within trial, as well as revealing what motivating forces drive those decision processes to update across trials.

In particular, much can be learned during circumstances of decision conflict. This can be operationalized in a number of ways. Especially if adopted within a neuroeconomic framework, multiple types of conflict scenarios that have not been easily measurable in the past using non-human animals can be constructed along a continuum of competing costs and rewards, including conflicts between reward and threat, between reward and self-inflicted pain, or between wanting highly desired although expensive rewards and knowing better to forgo such opportunities for economically smarter alternatives (Friedman et al. 2015; Guo et al. 2015; Resendez et al. 2016 Kim et al. 2017a, Sweis et al. 2018b; 2018c; Chapters 2 & 4).

Such neuroeconomic paradigms can pit multiple decision-making systems against one another and can reveal hidden information about the computational processes underlying specific aspects of behavior and within what neural constraints they might operate. Furthermore, such paradigms are naturalistic, do not rely on introspection to reveal dissociable cognitive processes, and are easily translatable across species (Sweis et al. 2018c; Chapter 4). Through this approach, computational processes of specific neural mechanisms can be revealed to be conserved across evolution and can be more thoroughly studied using the variety of tools, disease models, and patient populations available across species.

Off-line induction of circuit-specific plasticity

Many of the interrogation approaches used in the majority of recent optogenetic circuit-specific dissection studies in non-human animals rely on direct activity manipulations during on-going behaviors (Figure 6.3A-B) (Morales and Margolis 2017). That is, these manipulations are delivered “on-line” in real time affecting neural computations that drive behaviors. In fact, the vast majority of optogenetic behavioral studies generally fall under Figure 6.3A (Morales and Margolis 2017), while only a handful fall under Figure 6.3B (Schelp et al. 2017). In either case, two critical limitations exist in this approach in relation to studies of addiction and decision making. The first, previously discussed, is that the majority of behavioral tasks used

in circuit-dissection studies comprise simple tests of value. The second is that on-line manipulations impose disruptions of endogenous neural signaling and provide little insight into the functional consequences of synaptic remodeling on information encoding. Below, I describe a different way to think about experimental approaches to probe neural circuits that more directly links mechanisms of memory to decision-making processes in preparation for discoveries described in Chapter 9 of this thesis.

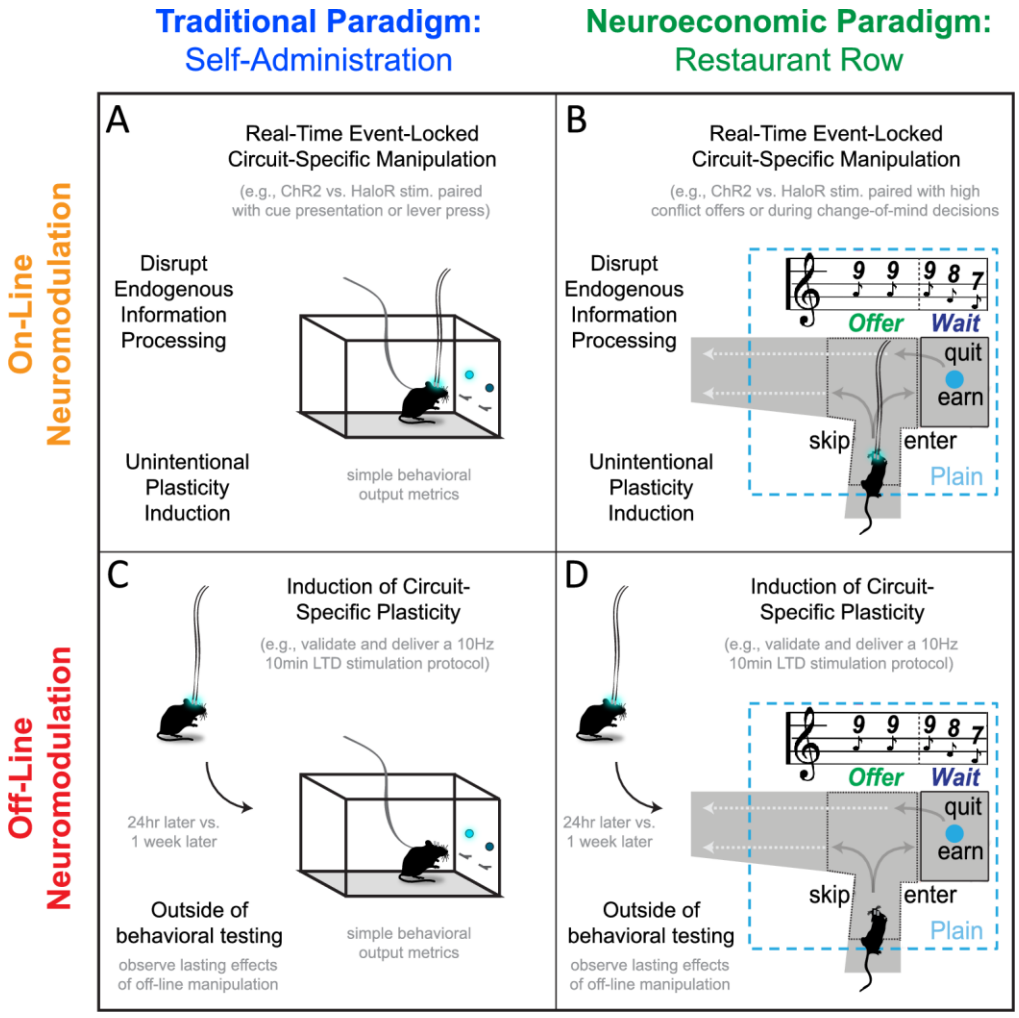
An alternative approach to neuromodulation interventions is to alter directly the synaptic efficacy of signal transmission in specific circuits through direct alterations in synaptic plasticity (Figure 6.3C-D). The goal of this approach is to change the weight of the information endogenously transmitted through a specific circuit, but not disrupt the information that is coded in this specific circuit during behavior. Importantly, these types of manipulations thus are delivered "off-line" outside of behavioral testing. To date, only a handful of studies fall under Figure 6.3C which I will discuss below while only a single recent study has adopted the approach in Figure 6.3D, which I will return to at the end of this chapter and present findings from in Chapter 9.

Only recently have tools been developed to directly manipulate the strength of circuit-specific synapses. Plasticity-altering interventions can now be delivered to specific synapses via optogenetics by enabling opsins in a circuit of interest and delivering a brief stimulation protocol in order to elicit lasting changes in synaptic efficacy. Using this approach, for example, excitatory opsins expressed in input-specific glutamatergic pathways can activated in a temporally precise manner to intentionally elicit long-term potentiation or depression. These tools have been used to good effect in several recent studies of the neurobiology of addiction (Pascoli et al. 2011; 2014, Ma et al. 2014, Hearing et al. 2016, Benneyworth et al. 2018). However, because such studies were performed using standard animal models of addiction (psychomotor sensitization, drug self-administration, or conditioned place preference), the functional consequences of synaptic remodeling on memory processes (e.g., consolidation, maintenance, extinction, or retrieval) and not decision-making processes were probed.

Studies of behavioral extinction in self-administration paradigms support a model of how plasticity in specific corticostriatal circuits are necessary for learned self-control (Peters et al. 2008, LaLumiere et al. 2010, Barker et al. 2012, Bossert et al. 2012, Gass and Chandler 2013, Keistler et al. 2015, Augur et al. 2016). For instance, long-term depression induced by optogenetically manipulating glutamatergic-specific excitatory pyramidal neurons that project from the infralimbic subregion of the prefrontal cortex to the
shell of the nucleus accumbens is capable of triggering reinstatement behaviors (Benneyworth et al. 2018). This idea is consistent with the "hypo-frontality" model of addiction that characterizes the inability of individuals with weaker corticostriatal connects to regulate maladaptive motivated behaviors (Kalivas and Volkow 2005; Bickel et al 2007; Chen et al. 2013; Camchong et al. 2014).

However, neuromodulation studies applying the same plasticity-inducing procedure in cocaine-abstinent mice vs. morphine-abstinent mice produce opposing findings using simple drug conditioned place preference tests (Hearing et al. 2016; Benneyworth et al. 2018). Similar optogenetically induced plasticity interventions (here, long-term depression) prevent drug-prime induced reinstatement of drug-seeking behavior in morphine-abstinent mice (Hearing et al. 2016), but provoke spontaneous reinstatement of drug-seeking behavior in cocaine-abstinent mice (Benneyworth et al. 2018). This distinction is critically important if we want to design treatments as it suggests the same treatment can be dysfunction-preventing in some situations, but dysfunction-provoking in others. This is especially concerning when addictions to different substances of abuse are often lumped together, both neuroscientifically and clinically. It is possible that cocaine-abstinent and morphine-abstinent mice may have undergone changes in fundamentally distinct computational processes despite appearing grossly similar by the end of extinction training (Badiani et al 2011). Thus, altering plasticity at a single connection can lead to drastically different behavioral outputs. By not carefully measuring the decision-making computational processes that may have separately gone awry in cocaine-abstinent mice vs. morphine-abstinent mice, knowing how to treat potentially distinct decision-making vulnerabilities becomes guess-work at best.

Figure 6.3: Neuromodulation intervention strategy in combination with task design matters



(A-B) Online neuromodulation manipulations (e.g., circuit-specific optogenetic stimulation) describe those where stimulation (either activation of excitatory opsins like channelrhodopsin-2 [ChR2] or inhibitory opsins like halorhodopsin [HaloR]) is delivered during on-going behaviors of interest. This could be time-locked to cue or lever-presentation in traditional paradigms (A) where extinction maintenance or reinstatement susceptibility can be assessed. This could also be time-locked to distinct decision-making action-selection processes in neuroeconomic paradigms (B) during re-evaluative change of mind decisions, for instance, only in high-conflict economic scenarios. However, in either (A) or (B), endogenous neural activity is disrupted. While this can reveal important information regarding on-going neural dynamics necessary or sufficient for certain behaviors, on-line neuromodulation actually reveals little regarding the functional consequences of synaptic plasticity in relation to addiction-related changes in neural circuitry. (C-D) Off-line neuromodulation interventions are capable of directly manipulating circuit-specific plasticity. For instance, well-characterized plasticity-inducing stimulation protocols (induction of long-term-depression in glutamatergic cortical pyramidal projections into the nucleus accumbens following 10min of 10Hz stimulation via ChR2) can be delivered acutely outside of behavioral testing. By observing lasting changes in behavior at later time points, the functional consequences of synaptic remodeling can be realized (e.g., mimicking disease states or reversing them). (C) By applying this approach in traditional paradigms, the functional consequences of circuit-specific synaptic remodeling on memory-related processes can be realized. (D) By applying this approach in neuroeconomic paradigms, the functional consequences of circuit-synaptic synaptic remodeling on separable decision-making computational processes can be realized.

Put simply, the functional consequences of synaptic remodeling on the discrete neural computations that drive on-going behavior remain ambiguous when tested in simple behavioral paradigms optimized for probing mechanisms of memory and not necessarily decision making.

Direct interrogation of the synaptic efficacy of specific circuits is an important tool for determining how information is processed as decisions are made. However, manipulating plasticity to understand the functional consequences of circuit-specific synaptic remodeling on distinct aspects of decision-making information processing is only as useful as the task utilized is sensitive to circuit-computation-specific behaviors. Combining circuit-specific off-line neuromodulation with complex testing that reveal separable behavioral computations is crucial. This combined approach will be critical for the development of disease-mitigating neuromodulation therapies, particularly those in which the benefit is intended to outlast the duration of neural stimulation and especially not unintentionally worsen disease prognosis. In order to begin to appreciate the complexity with which addiction-related processes can give rise to heterogenous circuit dysfunctions, next we will briefly discuss the varying plausible levels of addiction pathogenesis in this context before returning to how a combined approach of decision-making neuroeconomics and off-line plasticity manipulations can aid in resolving disease heterogeneity.

Next steps toward resolving addiction heterogeneity

Neuroeconomic tasks are capable of capturing scenarios of economic conflict that can operationalize decision-making concepts of reward value, self-control, and impulsivity in a number of different ways. The Restaurant Row task in rodents can measure the conflict between wanting highly desired rewards offers despite knowing better to seek out smarter alternatives in multiple decision-making valuation systems separated within the same trial (Sweis et al. 2018b, Chapter 2). This demonstrates a way for non-human animals to communicate "should vs. shouldn't" judgements through complex reward-seeking behaviors. This can model in rodents the difficult types of complex decisions humans recovering from addiction struggle with before relapsing. In the next chapter, I will present new findings that demonstrate how animal models of addiction can benefit from this approach and reveal hidden failure modes in separable decision processes

that are fundamentally distinct in different forms of addiction (i.e., cocaine vs. morphine). And in Chapter 9, I demonstrate how such lasting changes in behavior in only certain aspects of decision-making information processing can be causally linked to individual differences in circuit-specific mechanisms of memory. I will discuss the implications of these findings taken together that may suggest novel avenues for therapeutic interventions targeted to circuit-specific processes tailored toward an individual's computational dysfunction. I will also discuss the cautions revealed through these findings, where identical neuromodulatory interventions that may help one individual may in fact worsen disease state in another individual diagnosed with overtly similar behavioral dysfunctions.

Conclusion

Neural plasticity is purported to underlie long-lasting maladaptations in behavior, making addiction a chronic disorder of life-long struggle against relapse. It is this property of addiction – lasting changes in synaptic function that persist long after drugs of abuse have cleared an individual's system – that makes addiction a disorder of learning and memory and thus of decision-making. Recent advancements in decision science and neuroeconomics reveal that neurally dissociable valuation processes access distinct memories differently, and thus are uniquely susceptible to change as the brain changes during addiction. Therefore, addiction is best considered not as the disease itself but rather as a collection of symptoms of neurobiological diseases that are proving to be more heterogeneous than previously thought. Furthermore, addiction is not merely a disease of memory, but rather a consequence of changes in how decision-making processes access those memories. The advancement of our understanding of addiction etiology and thus development of better, lasting disease-augmenting therapies tailored to the individual will depend on tasks and neuromodulatory approaches that can access the complexity of circuit-computation-specific processes in the brain.

– Original Research –

Chapter 7

Prolonged abstinence from cocaine or morphine disrupts separable computation-specific valuations in the conflict between wanting and knowing better

Abstract

Relapse in recovering drug users is a key problem in addiction that is poorly understood. Recent theories suggest lasting changes in decision-making give rise to the life-long problems that drive relapse. These problems are rooted in continued use despite stated wishes not to. Such conflict, based on neuroeconomic theories of decision-making, is thought to arise from multiple valuation systems dependent on separable neural components. Much of the neurobiology of addiction literature however suggests that various drugs of abuse drive addiction through a unified mechanism. The majority of these studies are limited by the use of simple tests of value, leaving vulnerabilities in complex valuation processes and multiple etiologies of addiction largely unexplored. Here, I directly tested the effects of prolonged abstinence from different drugs in mice on a novel neuroeconomic task. This task reveals multiple tests of value and gains access to the operationalization of conflict between choosing smarter alternatives over highly desired rewards. Abstinence from repeated cocaine and morphine produced dissociable, long-lasting disruptions in separable decision-making processes of valuation conflict. Cocaine drove changes in deliberation before giving in to highly desired although economically unfavorable reward offers while morphine disrupted foraging re-evaluations after making decisions rapidly. These findings suggest that different drugs exploit distinct decision-making vulnerabilities (separate deliberation versus foraging processes) and that these changes occur at a timepoint relevant to relapse prevention. Our neuroeconomic approach to animal models of addiction can serve as a starting point in refining complex decision-making behavioral paradigms and highlights the need for computation-specific therapeutic interventions in separable decision-making processes to treat different forms of addiction. Enhancing sensitivity to identify multiple vulnerabilities in parallel processes that are otherwise masked by grossly similar maladaptations in behavior will aide in the development of treatments tailored to computation-specific neuropsychiatric disease endophenotypes.

Chapter reprinted with permissions from *Nature Communications*, modified from:

Sweis BM, Redish AD, Thomas MJ. 2018d. Prolonged abstinence from cocaine or morphine disrupts separable computation-specific valuations. *Nature Communications* 9: 2521.

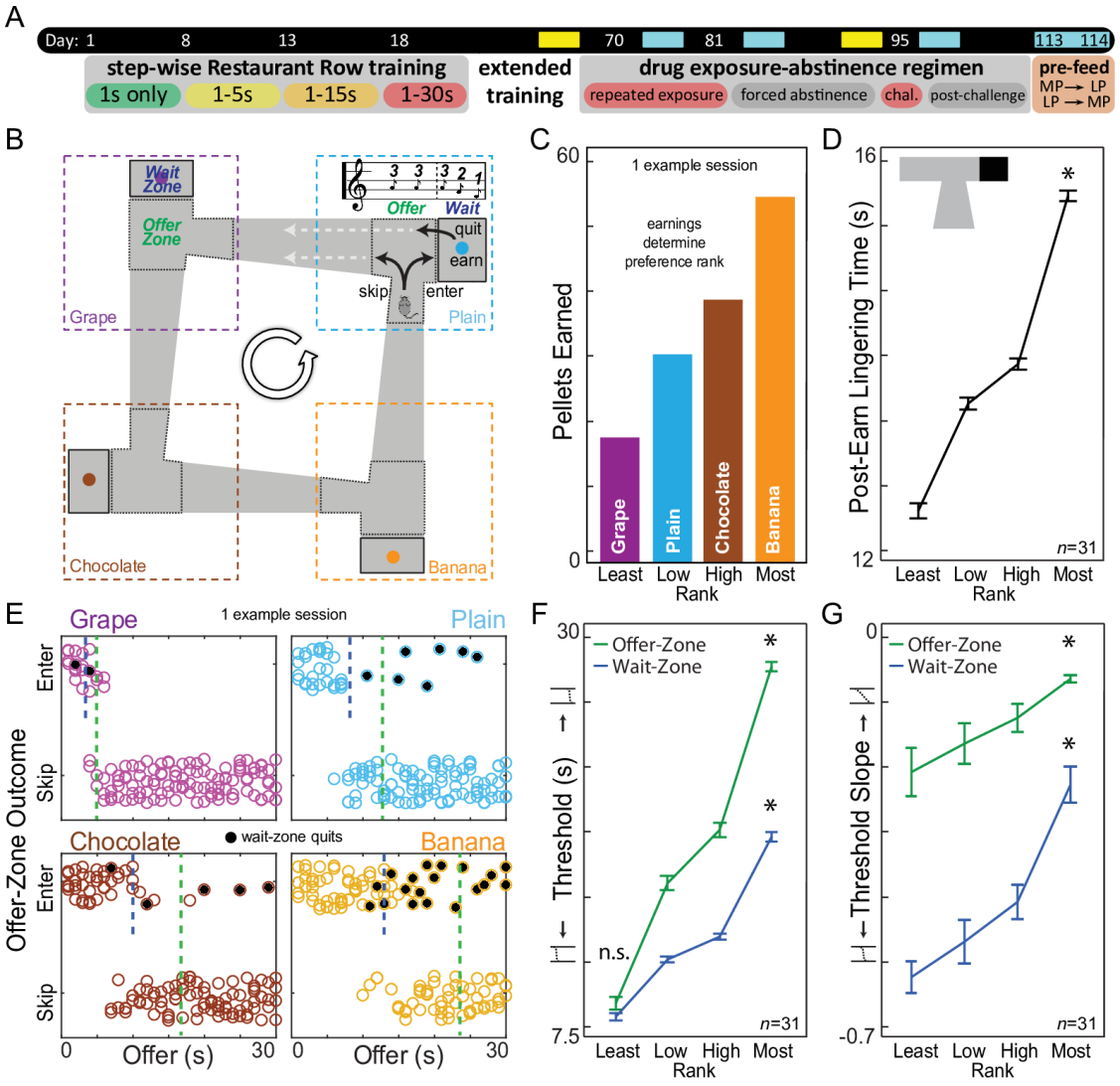
Introduction

Cocaine and morphine can both lead to rewiring of neural circuits involved in motivated behavior (Redish et al. 2008; Redish 2004). Although these drugs have different immediate mechanisms of action, theories have suggested that they ultimately converge on a final common dysfunction in mesolimbic dopamine leading to maladaptive reinforcement learning (Robinson and Berridge 2003; Nutt et al. 2015; Chiara 1999; Moal and Koob 2007; Thomas and Malenka 2003; Laviolette et al. 2004). However, it has also been hypothesized that malfunctions in decision-making systems with distinct neural circuits are capable of giving rise to multiple addiction endophenotypes, and that cocaine and morphine access different malfunctions in those circuits despite producing grossly similar changes in maladaptive goal-oriented behavior (Redish et al. 2008). So far, it has not been possible to dissect apart such changes behaviorally.

I developed a novel neuroeconomic task in mice that separates decision-making processes of reward conflict into behaviorally deconstructed stages (Steiner and Redish 2014). This task revealed multiple parallel valuation algorithms, dissociable processes of which were uniquely susceptible to different drugs of abuse.

Food-restricted mice traversed a square maze with four feeding sites (restaurants), each providing a different flavor, with two distinct zones: an offer-zone and a wait-zone (Figure 7.1B). Tones sounded upon offer-zone entry, whose pitch indicated a delay (pseudo-random, 1-30s) mice would have to wait if they chose to enter the wait-zone. Mice could choose to quit during delay countdowns. Importantly, mice had 1hr to forage for their food for the day. Using different flavors instead of pellet number allowed us to measure subjective preferences without introducing differences in feeding times.

Figure 7.1: Multiple valuations in Restaurant Row



(A) Experimental timeline. Timepoints of interest marked in yellow: well-trained at baseline (days 66-70, Fig.1-2); after prolonged abstinence from repeated drug exposure (days 90-94, Fig.3). Supplemental timepoints marked in cyan. (B) Mice were trained to run counter-clockwise around a square maze encountering serial offers for flavored rewards in four "restaurants." Tone pitch indicated delay length that sounded in the offer-zone, but did not countdown until after entering the wait-zone. (C) Flavors were ranked from least- to most-preferred by end-of-session earnings each day. Panel shows one example session. Mice showed individual preferences that were stable across days (Fig. S2). (D) Kruskal-Wallis tests revealed mice spent more time lingering at the reward site after earning rewards in more-preferred restaurants before moving on to the next trial (*P<0.0001). (E) Mice entered low delays and skipped high delays in the offer-zone, while infrequently quitting once in the wait-zone (black dots). Dashed vertical-lines represent calculated offer-zone and wait-zone "thresholds" of willingness to budget time. Green-line = offer-zone threshold (all offers). Blue-line = wait-zone threshold (entered offers). (F) KW tests revealed thresholds to enter (offer-zone) and earn (wait-zone) rewards were higher in more-preferred restaurants (*P<0.0001). Post-hoc Dunn's tests controlled for multiple comparisons revealed disparity between offer- and wait-zone thresholds was greater in more-preferred restaurants (*P<0.0001), generating more enter-then-quit events (least-preferred, not significant, n.s., P>0.05). (G) Slope of threshold fits were higher in the offer-zone than wait-zone and in more-preferred restaurants (KW test, *P<0.0001).

Methods

Mice

31-C57BL/J6 male mice, 13-weeks old, were trained in Restaurant Row. Mice were single-housed in a temperature- and humidity-controlled environment with a 12-hr-light/12-hr-dark cycle with water ad libitum. Mice were food restricted and trained to earn their entire day's food ration during their 1-hr Restaurant Row session. Experiments were approved by the University of Minnesota Institutional Animal Care and Use Committee. Mice were tested at the same time every day in a dim-lit room, were weighed before and after every testing session, and were fed a small post-session ration in a separate waiting-chamber on rare occasions to prevent extremely low weights according to IACUC standards (not <85% free-feeding weights).

Pellet training

Mice underwent 1-week of pellet training before being introduced to the Restaurant Row maze. During this period, mice were taken off of regular chow and introduced to a single daily serving of BioServ full nutrition 20mg pellets in excess (5g). This serving consisted of a mixture of chocolate-, banana-, grape-, and plain-flavored pellets. Next, mice (hungry, before being fed their daily ration) were introduced to the Restaurant Row maze 1 day prior to the start of training and were allowed to roam freely for 15min to explore, get comfortable with the maze, and familiarize themselves with the feeding sites. Restaurants were marked with unique spatial cues. Feeding bowls in each restaurant were filled with excess food on this introduction day.

Restaurant Row training

Task training was broken into 4 stages. Each daily session lasted for 1hr. At test start, one restaurant was randomly selected to be the starting restaurant where an offer was made if mice entered that restaurant's T-shaped offer-zone from the appropriate direction in a counter-clockwise manner. During the first stage (day 1-7), mice were trained for 1 week being given only 1s offers. Brief low pitch tones (4000Hz, 500ms) sounded upon entry into the offer-zone and repeated every second until mice skipped or until mice entered the wait-zone after which a pellet was dispensed. To discourage mice from leaving earned pellets uneaten, motorized

feeding bowls cleared any uneaten pellets upon wait-zone exit. Left over pellets were counted after each session and mice quickly learned to not leave the reward site without consuming earned pellets. The next restaurant in the counter-clockwise sequence was always and only the next available restaurant where an offer could be made such that mice learned to run laps encountering offers across all four restaurants in a fixed order serially in a single lap. Mice quickly learned not to run in the incorrect direction. During the second stage (day 8-12), mice were given offers that ranged from 1s to 5s (4000Hz to 5548Hz, in 387Hz steps) for 5 days. Offers were pseudo-randomly selected such that all 5 offer lengths were encountered in 5 consecutive trials before being re-shuffled, selected independently between restaurants. Again, offer tones repeated every second in the offer-zone indefinitely until either a skip or enter decision was made. In this stage and subsequent stages, in the wait-zone, 500ms tones descended in pitch every second by 387Hz steps counting down to pellet delivery. If the wait-zone was exited at any point during the countdown, the tone ceased and the trial ended, forcing mice to proceed to the next restaurant. Stage 3 (day 13-17) consisted of offers from 1s to 15s (4000Hz to 9418Hz) for another 5 days. Stage 4 (day 18-70) offers ranged from 1s to 30s (4000Hz to 15223Hz) and lasted until mice showed stable economic behaviors (Supplementary Video). We used 4 Audiotek tweeters positioned next to each restaurant powered by Lepy amplifiers to play local tones at 70dB in each restaurant. We recorded speaker quality to verify frequency playback fidelity. We used Med Associates 20mg feeder pellet dispensers and 3D-printed feeding bowl receptacles fashioned with mini-servos to control automated clearance of uneaten pellets. Animal tracking, task programming, and maze operation was powered by AnyMaze (Stoelting).

Restaurant Row metrics

Vicarious trial and error behavior (VTE) was measured as the absolute integrated angular velocity of a mouse's X-Y-position over the course of time and distance from tone-onset upon entry into the offer-zone until exiting the offer-zone (either toward the wait-zone or toward the corridor heading to the next restaurant). VTE of a given path-trajectory was measured in IdPhi units. Reaction time in the offer-zone was also measured in this period.

Reaction time to quit was also measured in the wait-zone from tone-count-down-onset until exit from the wait-zone prematurely before a pellet is earned. Post-earn-consumption-and-lingering-time was measured from pellet delivery-onset until the first exit was made out of wait-zone. In an earlier pilot study, cameras were placed in the wait-zone in order to observe "lingering" behaviors. After immediate pellet consumption, mice exhibited no unusual behaviors other than occasional grooming and checking the empty pellet receptacle for varying lengths of time before exiting and proceeding to the next restaurant.

Offer- and wait-zone thresholds were measured for each session by fitting sigmoid functions to zone choice outcomes as a function of offer delay, restaurant by restaurant. Inflection point and slope of each sigmoid fit was calculated. In order to calculate the value of the offer on any given trial, thresholds were re-calculated in a "leave-one-out" analysis excluding the current trial. We then used wait-zone threshold minus offer to calculate value.

Economic conflict inefficiency was measured both for the offer-zone and wait-zone. This metric characterized how mice responded to an economically unfavorable offer (an offer where the delay was greater than wait-zone threshold). The ratio of the probability of entering the wait-zone for offers above the wait-zone threshold relative to skipping them was calculated in each restaurant as a function of rank. Similarly, in the wait-zone, after mice had already accepted such offers greater than wait-zone threshold, we characterized how long it took an animal to quit such an offer. If mice took so long that the amount of time remaining when quitting was less than wait-zone threshold, that was characterized as an economically inefficient quit. The ratio of the probability of quitting these offers after they counted down passed wait-zone thresholds relative to quitting before the countdown passed wait-zone thresholds was calculated in each restaurant as a function of rank.

In order to control for the possibility that the analysis of changes in VTE in the offer-zone in economically unfavorable acceptances (taking offer-zone deals that are above the wait-zone thresholds) could have been affected by unequal or different distributions of offers based on trial type (e.g., skipping offers, entering offers

above threshold, or entering offers below threshold), we generated simulated shuffled data sets of reaction time and VTE when both skipping and entering offers below threshold matching the same trial-by-trial distributions of offer lengths as those subsets of trials where mice entered offers above threshold. This ensures any changes seen in offer-zone behaviors, particularly when entering economically favorable vs. unfavorable offers, are not skewed by differences in distribution of trials of different offer lengths.

Modeling

Economic deliberative algorithm models were generated via Matlab simulations where we calculated the probability of entering vs. skipping offers as a function of increasing delays from 1 to 30s of two offers (the current offer (d1), and the expected next offer (d2)). The shape of this value function ($V = 1 / (1+k*d1) - 1 / (1+k*d2)$) changes with increasing k (increasing impulsively hyperbolic functions).

Drug exposure regimen and locomotor sensitization

After ~70 days of training mice were injected with saline (0.9% NaCl) for three days in order to get them acclimated to the stress of injections. All injections were volume-corrected after measuring mouse body weights right before injections. Next, mice received 12 evenings of repeated drug or saline control injections. This is a standard and well-established drug-treatment regimen known to produce robust and long-lasting drug-related changes at a stage just before relapse. Overall, our goal was to measure how decision processes were affected by repeated drug use, rather than acutely when animals were on drug. Thus, it is the prolonged abstinence timepoint ~2 weeks following the 12th drug injection that is of importance.

This drug treatment regimen is a simple, straightforward yet powerful means of producing robust and long-lasting behavioral and neurobiological changes linked to aspects of addiction such as incentive sensitization and neural plasticity in the mesocorticolimbic dopamine system. By looking at a time point during prolonged abstinence, we intended to characterize changes that may reflect the life-long decision-making problems seen in recovering addicts. Long-lasting forms of neurobiological plasticity changes are observed at these

prolonged abstinence time points coinciding with and causally linked to escalation of craving. Such plasticity measurements predict relapse susceptibility in human addicts.

Injections took place in the evening 4 hours post-Restaurant Row testing. Our goal was to expose animals to drugs of abuse outside of testing hours, to be especially sure drug has cleared the animals' system before the next day's behavior. Furthermore, we wanted to avoid the effects of acute withdrawal on each day of Restaurant Row testing during the drug exposure phase. Repeated Restaurant Row testing during the drug exposure phase was not intended to capture instances when drug is on board, nor was it intended to compare changes between first and subsequent drug exposures, nor was it intended to analyze the effects of immediate cessation of repeated drug administration on decision-making. Instead, the goal was to interrogate decision-making after prolonged abstinence. Repeated Restaurant Row testing during the drug exposure phase and early abstinence was mainly intended to (1) ensure the animals did not "unlearn" the task day to day, and (2) maintain regular self-earned food-intake amounts contingent upon task performance rather than giving the animals non-contingent food or "days off."

In the evening at the time of each drug injection, mice were placed in large locomotion monitoring boxes with tracking cameras fixed above automatically measuring distance traveled using AnyMaze software (Stoelting). Mice were placed in the boxes for 20min before being injected intraperitoneally with saline and then monitored for 90min post-injections. Then mice were divided into three groups: saline (n=10), cocaine (n=10), and morphine (n=10). One mouse out of the original 31 was excluded because it never learned the task. Mice were then injected with their respective treatment for 12 consecutive nights while being tested in Restaurant Row regularly. For the drug groups, mice were given lower doses (15mg/kg cocaine, 10mg/kg morphine) on the first and last nights and received repeated higher doses (30mg/kg cocaine, 20mg/kg morphine) on the intermediate 10 nights. Three mice were lost during the drug phase in the cocaine group and were excluded from analyses. Mice were then put through a forced abstinence period for 2 weeks while regularly being tested in Restaurant Row.

In addition to the prolonged abstinence timepoint that is the main focus of the drug paradigm, we also introduced animals to an acute drug challenge at the end of the ~2 weeks of abstinence timepoint. This was intended to probe responsivity to a drug prime and assess degree of locomotor sensitization that typically incubates over prolonged abstinence and can be expressed upon drug-re-exposure. Locomotor sensitization was measured as the psychomotor response measured immediately following drug injection at this timepoint compared to psychomotor response measured immediately following drug injection on the 12th evening of the repeated drug exposure sequence. We randomly injected mice 3 times with saline across the evenings before experiencing this acute drug challenge, again, to acclimate the animals to the stress of injections in preparation for the forthcoming drug-re-exposure challenge.

Mice were challenged in the evening with a single low dose of drug same dose as the 1st and 12th night of drug in the repeated drug exposure sequence, being re-exposed to the same drug administered previously. Saline mice were divided into two groups of n=5 to receive a low dose of either cocaine or morphine for the first time, acutely. Despite the small sample size, this split was done to ensure that sensitized locomotion in response to a single dose was present only in animals with a history of repeated drug exposure. This comparison was statistically significant even with samples of n=5. This replicates work from our lab and numerous others. Regardless, the primary analyses (comparing baseline to prolonged abstinence) occurred before the saline group was split and statistics were done with the complete saline group as control.

Following the acute drug-re-exposure challenge, Restaurant Row was tested regularly during the day for an additional 2-3 weeks. Because there were no lasting drug effects on any animal behavior in the formerly saline animals after the acute drug-re-exposure challenge session which took place ~20 days before the pre-feeding probe sessions (described below), this group served as “control” conditions for the pre-feeding probe sessions.

Devaluation / invigoration pre-feeding probe sessions

The pre-feeding probe sessions were performed at the end of the experiment and were intended to elucidate if rapid decisions or "snap-judgments" were flexible or inflexible processes. Devaluation probes are often used to differentiate goal-oriented (flexible and thus sensitive to devaluation) and habitual (inflexible and thus insensitive to devaluation) decision processes. The devaluation probe in our task allowed us to rule-out habitual processes. There was no further testing after the pre-feeding probes as the experiment ended and all mice were retired.

Mice were pre-fed 30-60min before testing in an amount equivalent to what they typically earned in their most-preferred restaurant. Since each animal showed individual revealed preferences (i.e. different animals like different flavors best), we fed each animal its most-preferred flavor on one day and its least-preferred on the next. Since some animals received their most-preferred flavor on the first-day of pre-feeding while others received their least-preferred flavor on the first-day of pre-feeding (randomly selected and counter-balanced), day two of pre-feeding flipped this assignment. There were no order effects and no lasting body weight changes on day one versus day two of pre-feeding, so we pooled together the first and second day of pre-feeding to look at group differences between being fed one's most-preferred flavor versus least-preferred flavor.

The fact that all groups still showed sensitivity to the pre-feeding probe (although with intricate fine-grained differences between groups described in the Supplementary Discussion), we determined that the decision-processes in Restaurant Row remained flexible and had not transitioned to habit-like processes.

Statistical analyses

All statistical analyses were carried out using JMP Pro 13 Statistical Discovery software package from SAS. Statistical significance was assessed using non-parametric statistical tests, as the data was not normally distributed (offer-zone time, offer-zone VTE, wait-zone quit time, post-earn linger time, and offer- and wait-zone thresholds all reject normal distributions using the Kolmogorov-Smirnov-Lilliefors test for goodness of

fit, P<0.01). Described below are the statistics used for each main figure, where applicable. Statistics for Supplementary Data are detailed in corresponding figure captions or in the Supplementary Discussion. All error bars are expressed as ± 1 SEM. Asterisks used in figures are intended to direct attention to comparisons of interest. See below in this section/Supplementary Data captions/Supplementary Discussion for tests used and significance thresholds.

Figures 7.1A-C,E, 7.4A-BE,H-I, 7.9A-B, 7.15A-I are illustrative in nature, single session examples, or intended to demonstrate derivation of a higher-order metric summarized for comparison in a separate figure, and thus, analyses reports are deemed not appropriate or not included.

The Kruskal-Wallis (K-W) test was used as a non-parametric equivalent to the parametric one-way ANOVA test in Fig. 7.1D,F-G, 7.4C-D,F-G,M-N to test dependent measures against flavor rankings (or against the three conditions described in 7.4L). Post-hoc analyses controlling for multiple comparisons were performed using Dunn's test to preserve pooled variance from the K-W test in order to compare conditions in a pairwise manner. Much of these comparisons included testing flavor rankings pairwise (e.g., most-preferred to least-preferred) as well as to compare values of the same flavor ranking across levels of an separate factor stated on each figure (e.g., skip vs. enter, offer-zone vs. wait-zone). K-W tests were significant across rank on all metrics in the above figures (P<0.0001) except in Fig.7.4C-D for the enter condition (P>0.05). Dunn's tests showed that the most-preferred flavor was significantly greater than the least-preferred flavor on all metrics in the above figures (*P<0.0001). Dunn's test also showed that offer-zone thresholds and slope were greater than wait-zone thresholds and slope (Fig.7.1F-G, *P<0.0001), except between thresholds in least-preferred restaurants (Fig.7.1F, P>0.05). Dunn's test also showed that skips were greater than enters in both offer-zone time and VTE in all restaurants (Fig.7.4C-D, *P<0.0001). Lastly, K-W and Dunn's tests on quitting behavior in Fig.7.4L confirm economically inefficient quits made up the majority of quit events in the wait-zone (*P<0.0001).

In addition to the significant interactions across rank in Fig.7.4F,N, the Sign test was used to assess if behavior in each restaurant was above or below the 1:1 ratio line on economic inefficiency in the offer-zone (Fig.7.4F) and the wait-zone (Fig.7.4N). Data above the 1:1 ratio line, or a positive sign, indicates economically inefficient behavior. Only behavior in the offer-zone of the most-preferred flavor was above the 1:1 ratio line (Fig.7.4F, $P<0.0001$), and not for other flavors in the offer-zone nor any flavor in the wait-zone (Fig.7.4N, $P>0.05$).

The Kolmogorov-Smirnov test was used to assess differences in cumulative probability distributions of offer-zone time and VTE in Fig.7.4J-K, 7.9C-D. Our comparison of interest was between enters for offers above wait-zone threshold and enters for offers below wait-zone threshold, which at baseline were not statistically different from each other in both time and VTE (Fig.7.4J-K, $P>0.05$). This was replicated at the prolonged abstinence timepoint in both the saline and morphine groups ($P>0.05$), but not cocaine group ($*P<0.01$) for both offer-zone time and VTE (Fig.7.9C-D).

Correlations in Fig.7.9E were performed on the degree of change in locomotor sensitization and the degree of change in the ratio of offer-zone times between economically inefficient enters and economically efficient enters across timepoints (comparing prolonged abstinence to baseline). At baseline, the ratio of enter times between these two economic scenarios is 1:1 (both types of enters are indistinguishable snap-judgments at baseline). Only in the cocaine group, this metric increases above a 1:1 ratio (enter times for economically unfavorable offers are longer). Thus, the change in this ratio from prolonged abstinence to baseline in the cocaine group is greater than zero, while remaining at zero for the morphine group. Pearson correlation coefficients was 0.685 in the cocaine group ($*P<0.1$) and -0.078 in the morphine group ($P>0.1$). Given the small sample size of the cocaine group ($n=7$), the correlation coefficient of the cocaine group lies within the 90%, but not 95% confidence interval.

The Friedman test was used as a non-parametric equivalent to the parametric one-way ANOVA with repeated measures in Fig.7.9F when comparing thresholds across two timepoints (baseline and prolonged abstinence).

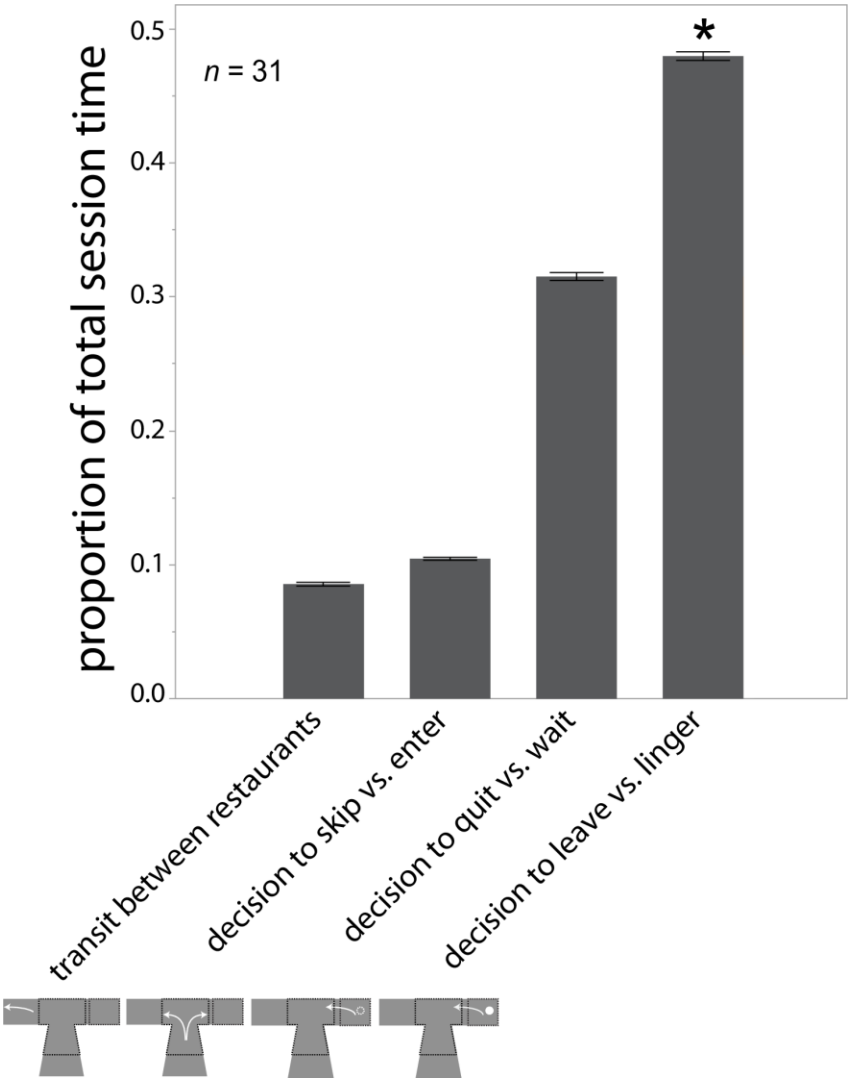
Only in the morphine group did wait-zone thresholds significantly increase across time points (*P<0.05) while offer-zone thresholds did not, nor either threshold in the saline and cocaine groups (P>0.05). Post-hoc analyses using Mann-Whitney tests while correcting for multiple comparisons allowed for non-parametric comparisons at either timepoint between offer-zone and wait-zone or between drug conditions. At the prolonged abstinence timepoint, in the morphine group, wait-zone thresholds were significantly higher than offer-zone thresholds (*P<0.05), which were no different at baseline or at either timepoint in the saline and cocaine groups (P>0.05). Lastly, wait-zone thresholds at the prolonged abstinence timepoint in the morphine group was significantly higher than the saline group (*P<0.05) while comparison of wait-zone thresholds between cocaine and saline animals were no different at the prolonged abstinence timepoint (P>0.05).

Results

The economic key to foraging is the division of time. Time spent choosing in the offer-zone, waiting in the wait-zone, and remaining at the reward site after receiving food all detracts from time spent making other decisions elsewhere. Critically, choices in each of these three decision modalities (skip vs. enter, quit vs. continue to wait, leave vs. linger) are computationally distinct valuation processes.

Mice spent the majority of time lingering at the reward site after earning and consuming a reward (Figure 7.2) Interestingly, mice lingered longer in more-preferred restaurants (Figure 7.1D). This decision to linger rather than leave, where no overt reward is being sought out, may represent a conditioned-place-preference-like effect associated with each restaurant's context.(Clark et al. 2012)

Figure 7.2: Allocation of total session time budget across multiple separate valuation behaviors



(Left to right) Average percent of total session time spent traveling between restaurants, deliberating in the offer-zone (skipping vs. entering, measured between initial tone onset and offer-zone exit), foraging in the wait-zone (investing time before quitting vs. earning pellets, measured between tone countdown onset and premature wait-zone exit or pellet delivery), and consuming food and lingering at the reward-site after earning pellets (measured from time of pellet delivery to wait-zone exit). Majority of total session time was spent lingering at the reward site compared to other task behaviors (Friedman, $P<0.0001$, post-hoc Mann-Whitney comparisons against lingering time, $^*P<0.0001$).

We calculated offer-zone "thresholds" of willingness to enter as a function of offered delay (Figure 7.1E, Figure 7.3), and found higher thresholds in more-preferred restaurants compared to less-preferred restaurants (Figure 7.1E-F). Interestingly, mice took longer in the offer-zone deciding to skip than deciding to enter (Figure 7.4A-C). Furthermore, decision time took longer when skipping more-preferred restaurants (Figure 7.4C). These data suggest that highly desired rewards were more difficult to turn down.

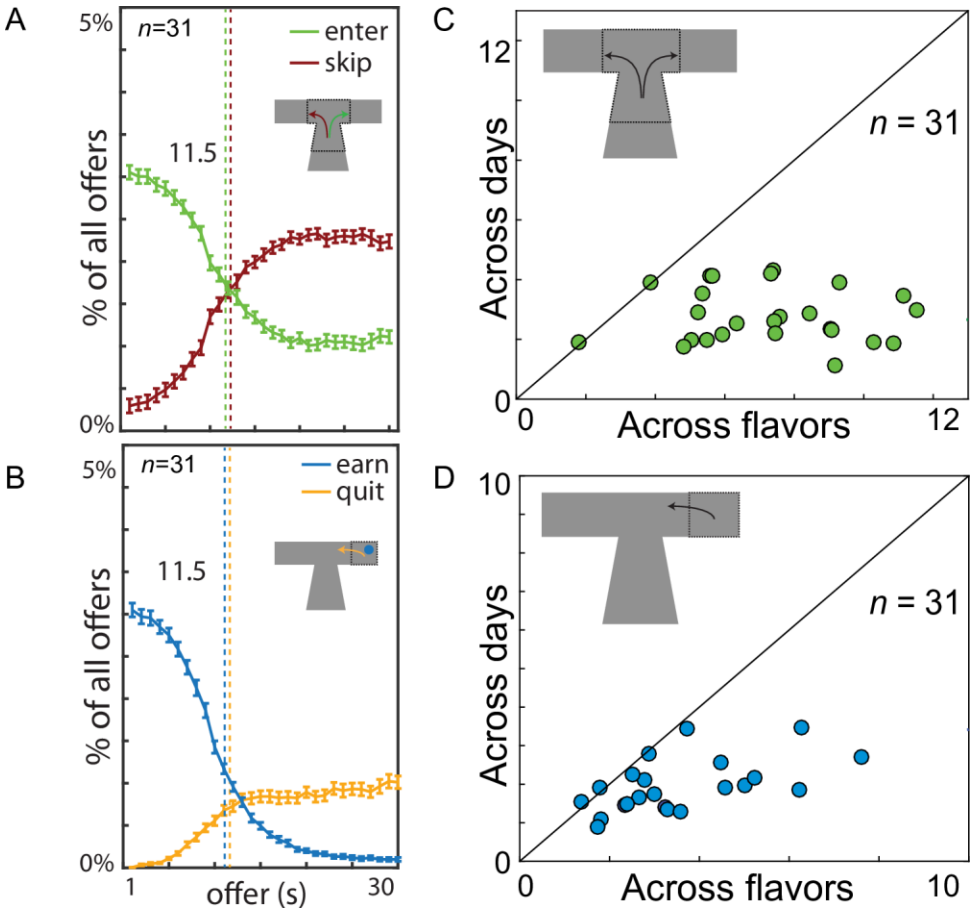
Degree of adherence to thresholds can be measured via slope of fitted sigmoid functions. Steeper (more negative) slopes indicate low likelihoods of threshold violation (i.e., enter above or skip below offer-zone thresholds). Threshold slope was less steep in more-preferred restaurants (Figure 7.1G), suggesting highly desired reward offers blurred subjective policy judgments.

We carried out similar analyses in the wait-zone for quit decisions. Wait-zone thresholds also increased for more-preferred flavors (Figure 7.1E-F). However, wait-zone threshold slope was steeper than offer-zone threshold slope (Figure 7.1G), indicating mice were less likely to violate wait-zone thresholds. This meant that wait-zone metrics captured a fundamentally different valuation process than the offer-zone.

Disparity between offer- and wait-zone thresholds was greatest (offer-zone>wait-zone) in more-preferred restaurants (Figure 7.1F). In these restaurants, then, mice were more likely to accept offers with a higher cost than subjective value indicated that they should (Figure 7.4F). This scenario – entering offers that are greater than wait-zone thresholds – is an explicit economic failure to choose a better alternative over a tantalizing reward offer. In such instances, it would have been economically advantageous to choose to skip in the offer-zone.

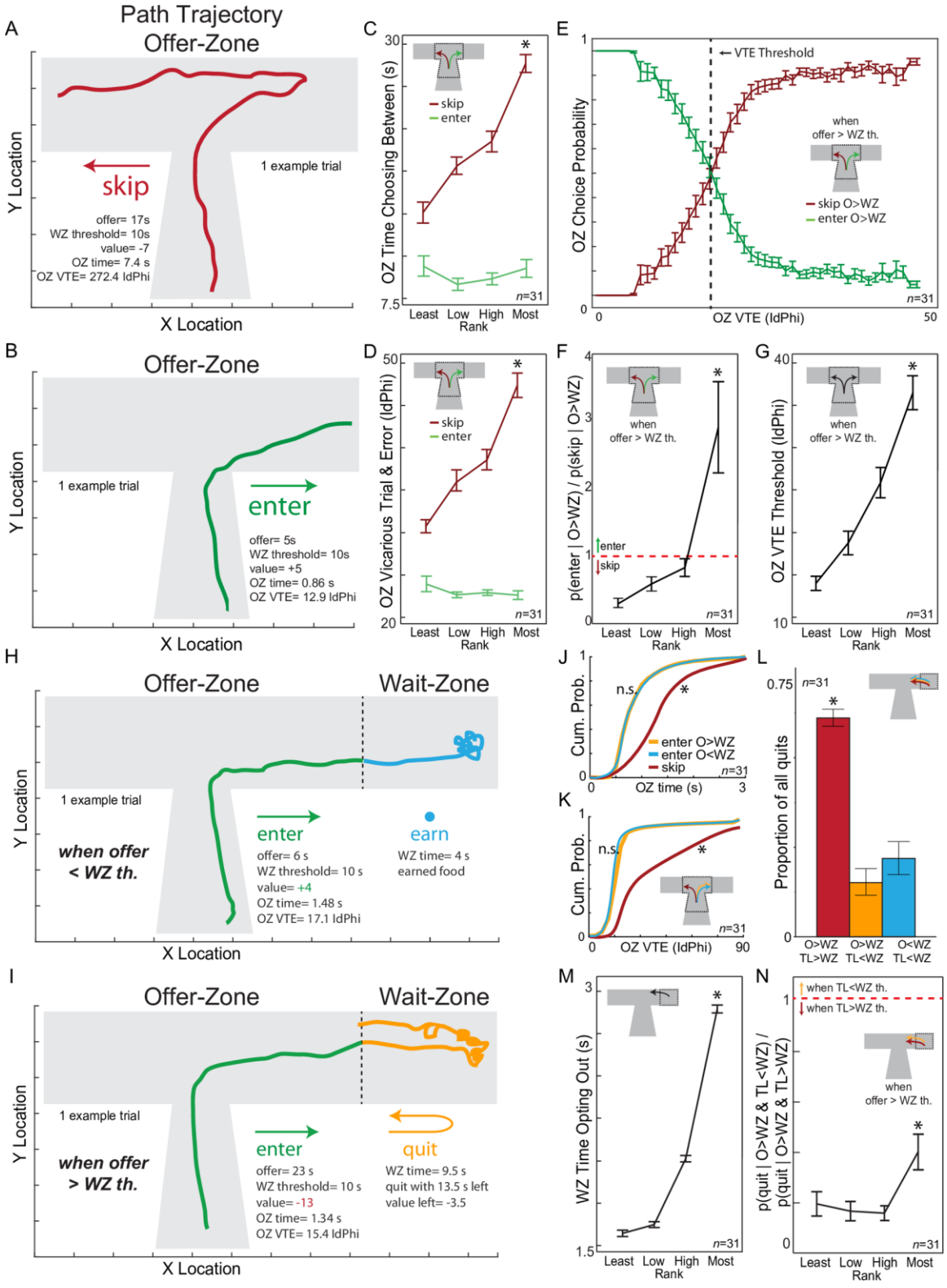
Because path-trajectories can reveal decision-making processes, we examined moment-by moment body positions during offer-zone decisions (Redish 2016; Muenzinger 1956; Tolman 1939). We found that mice often oriented first toward entering the wait-zone before pausing, re-orienting, and then ultimately deciding to skip (Figure 7.4A-B). This behavior is a well-studied decision-making phenomenon termed vicarious

Figure 7.3: Offer discrimination and threshold stability



(A) In the offer-zone, mice accepted (entered) short offers while skipping long offers. (B) In the wait-zone, mice waited for (earned) short offers while quitting long offers. (A-B) Vertical dashed-lines indicates overall threshold collapsed across restaurants (~11.5s in both zones). (C-D) Variability of offer-zone thresholds (C) and wait-zone thresholds (D) was calculated between flavors (x-axis) as well as for a given flavor across 10 days of stable performance (y-axis). Dots represent individual subjects. Space below unity line reflects range of idiosyncratic variability in individual differences in subjective flavor preferences while also reflecting stable preferences within flavor (low relative variability).

Figure 7.4: Separating deliberation and foraging conflicts between wanting and knowing better



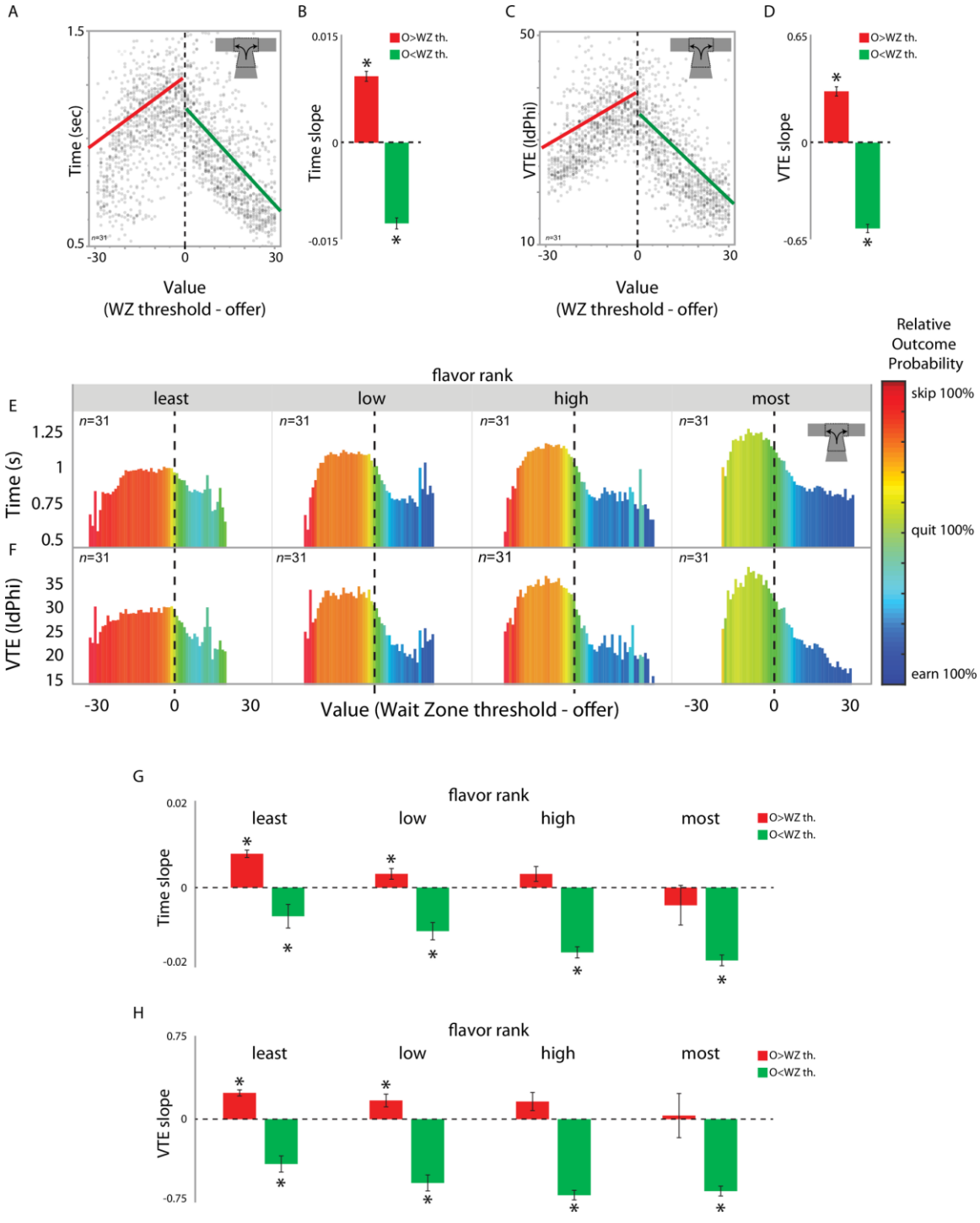
(A-B) Example x-y-locations of a mouse's path-trajectory in the offer-zone over time during a single trial. (A) Skip decision for a high delay offer. The mouse initially oriented toward entering (right) then ultimately re-oriented to skip (left). Wait-zone threshold minus offer captures the relative subjective "value" of the offer. Negative value denotes an economically unfavorable offer. (B) Enter decision for positively valued offer; rapid without re-orientations. Reaction time (C) and VTE (D) behavior was higher for skip compared to enter decisions and only increased in more-preferred restaurants for skip decisions (KW tests, *P<0.0001). (E) Mice were more likely to skip negatively valued offers the more they displayed VTE behavior. Vertical dashed-line indicates the amount of VTE required to skip these offers 50% of the time. (F) Mice were more likely to enter these offers in higher-preferred restaurants, entering more than skipping in only most-preferred restaurant (KW and Sign tests, *P<0.0001). (G) Amount of VTE required to reliably skip these offers was higher in more-preferred restaurants (KW tests, *P<0.0001). (H-I) Example path-trajectory in the offer- and wait-zones. (H) Rapidly entering then earning a positively valued offer. (I) Rapidly entering then quitting a negatively valued offer. (J-K) Cumulative probability distribution of offer-zone time (J) and VTE (K) for skips and enters split by offer value. Both types of enter decisions were rapid compared to skips (Kolmogorov-Smirnov tests, *P<0.05) and indistinguishable from each other (KS tests, not significant, n.s., P>0.05). (L-M) Majority of quits took place for negatively valued offers and while time left was still greater than wait-zone thresholds (L), despite taking longer to quit in more-preferred restaurants (M, KW-D tests, *P<0.0001). (N) Although mice were more likely to quit negatively valued offers while the amount of time left was still above wait-zone thresholds in all restaurants, they were less capable of doing so in more-preferred restaurants (KW and Sign tests, *P<0.0001).

trial and error (VTE) that reveals on-going deliberation and planning during moments of indecision (Supplementary Discussion, Tolman 1939; Muenzinger 1956; Redish 2016). We measured VTE as the absolute integrated angular velocity over the course of a given path-trajectory (IdPhi). There was more VTE (IdPhi was larger) during skip decisions in general and particularly so when skipping in more-preferred restaurants (Figure 7.4A,D, Figure 7.5). The presence of VTE suggests that in the offer-zone, decisions to skip included a delayed valuation that overrode initial rapid decisions. This provides a potential point of decision-making vulnerability in addiction – one rooted in failure of a deliberative or planning process when engaged in conflict between a highly desirable reward vs. choosing smarter alternatives. We know that VTE is a sign of deliberation but VTE has not yet been measured in an addiction model.

Interestingly, skipping offers above wait-zone thresholds was more likely to occur the more an animal displayed VTE behavior (Figure 7.4E). This suggests that the more a planning process was engaged, the less likely desired rewards could out-compete making smarter choices, independent of offer value (Figure 7.6). By classifying the amount of VTE required to skip these economic scenarios at least 50% of the time, we found that skipping high delays in more-preferred restaurants required greater amounts of VTE (Figure 7.4G). Furthermore, we found enters for offers above versus below wait-thresholds were both rapid and indistinguishable in reaction time and VTE (Figure 7.4H-K), suggesting reward-taking behaviors were generally snap-judgments while reward-opposing behaviors were not.

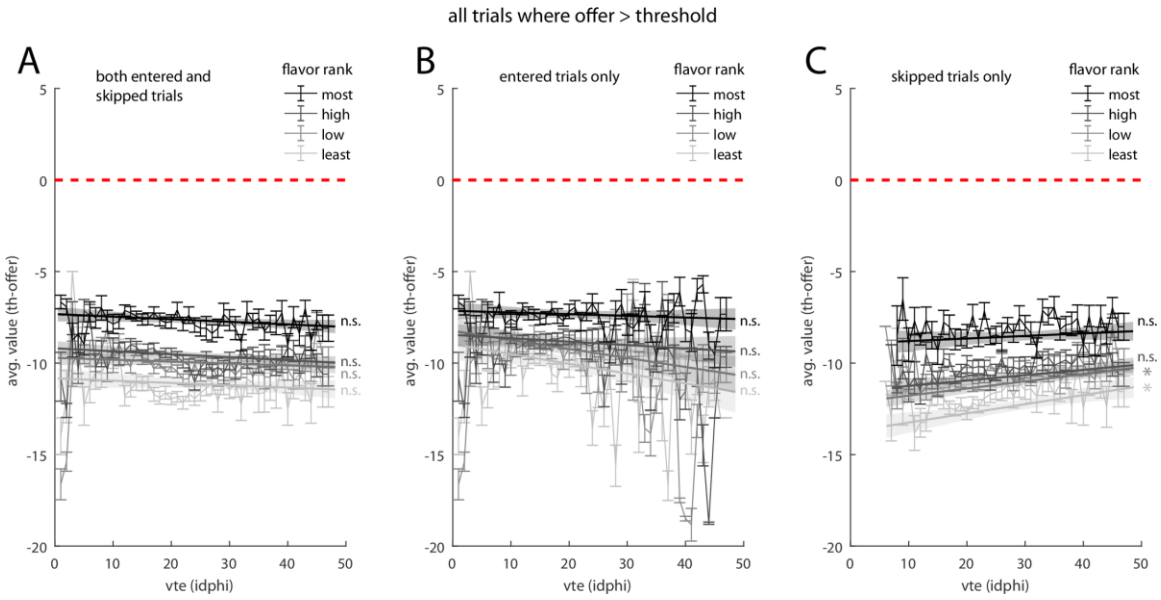
As noted, mice were more likely to err by entering offers above wait-zone threshold in more- vs. less-preferred restaurants (Figure 7.4F). In the wait-zone, mice were more likely to quit after enters above than after enters below wait-zone threshold. Moreover, they were more likely to quit while the amount of countdown time left remaining was still above the wait-zone threshold (Figure 7.4L, Figure 7.7). Thus, wait-zone decisions to quit were advantageous change-of-mind re-evaluations correcting economically unfavorable rapid valuations made in the offer-zone. This reveals that mice, despite making economically unfavorable decisions in the offer-zone, could remediate those initial snap-judgments.

Figure 7.5: Offer zone deliberation behaviors distributions by value, rank, and trial outcome



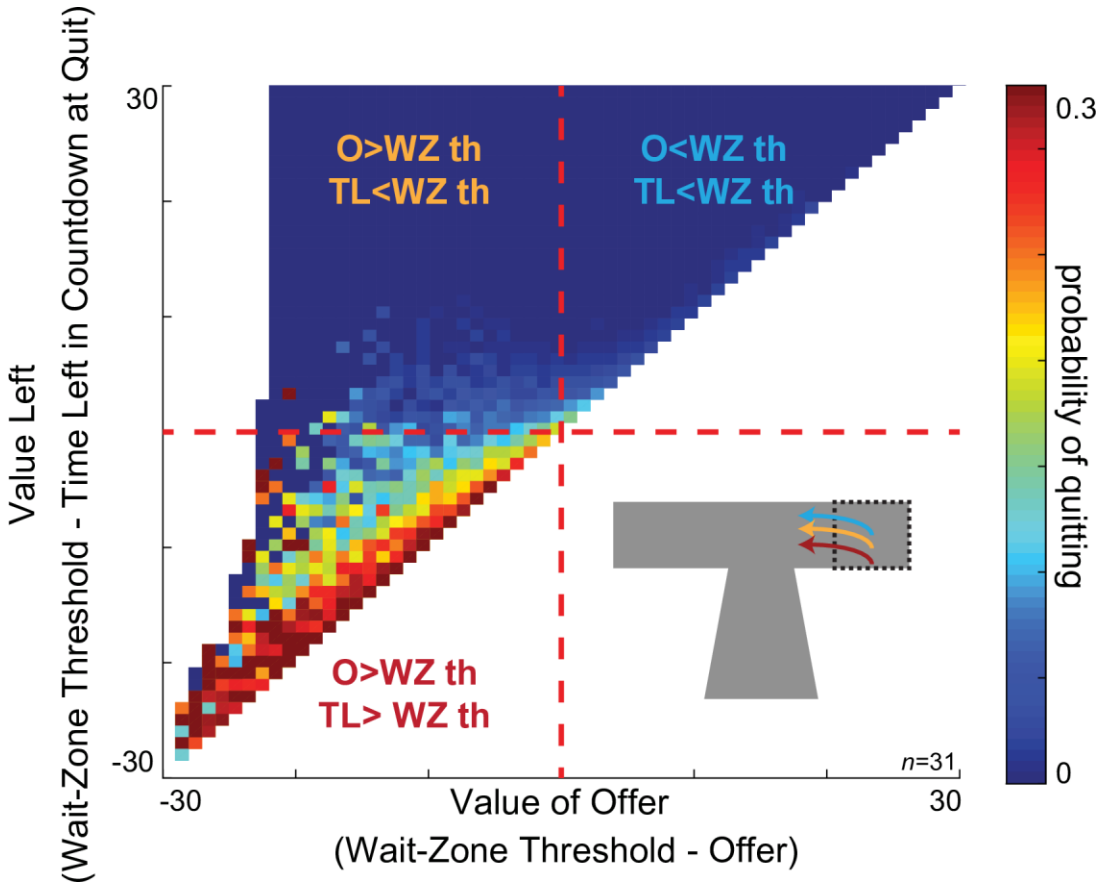
(A-D) Linear fits for offer-zone time (A-B) and VTE (C-D) as a function of wait-zone-threshold-derived value revealed decisions got progressively easier (less time and less VTE) when offers were farther away from threshold in either direction (B,D, Sign test slope is significantly different from zero, *P<0.05). That is, offers near threshold (zero value, vertical dashed black line) were toughest. (E-H) Same as A-D split by subjective flavor preference rankings. (E-F) Color scale describes the relative likelihood a trial at a given wait-zone-derived value is to end as either a skip, quit, or earn outcome. Note the increasingly sharper leftward color transition toward red (reflecting skip events) for negatively valued offers in less-preferred restaurants compared to the broader leftward color transition that is predominately green for negatively valued offers in more-preferred restaurants (reflecting enter-then-quit events). (G-H) Slopes for both time (G) and VTE (H) for positively valued offers are significantly different (less) than zero in all ranks, while slopes for negatively valued offers only in less-preferred restaurants are significantly different (greater) than zero. This indicates that decisions for worse deals in more-preferred restaurants, unlike in less-preferred restaurants, were not any easier to make. (*P<0.05).

Figure 7.6: Controlling for value as a function of vicarious trial and error (VTE)



The average value of offers encountered as a function of VTE measured on that trial are plotted split by restaurant ranking as well as decision outcome on that trial (A: skips and enters grouped together, B: enters only, C: skips only). Data presented here are derived from trials where offer > wait zone thresholds, representing "bad" deal trials. Thus, average values for all offers plotted here are <0 (horizontal dashed red line). As a function of VTE, offer value could explain some but not all changes in VTE. Correlation significance controlling for 12 multiple comparisons, Bonferroni corrected alpha level 0.05, *P<0.004, not significant (n.s.) P>0.004.

Figure 7.7: Economic efficiency of quit events in the wait-zone



Accepted offers greater than wait-zone threshold are negatively valued offers (wait-zone threshold greater than offer). This is separated on the x-axis to the left of the vertical dashed red line indicating zero value (offers at wait-zone threshold). Additionally, the time remaining in the countdown at the time of quit was measured and "value left" was calculated by subtracting wait-zone threshold minus time left. Thus, negative value left in the countdown at quit is separated on the y-axis below the horizontal dashed red line indicating zero value left (time left in countdown at quit at wait-zone threshold). Majority of quits took place in the lower left quadrant (summarized in Fig.2L), indicating that the majority of quits occurred after mice had taken offers greater than their typical threshold (i.e. economically unfavorable), and the quit was a form of self-correction.

We found that mice took longer to quit in more-preferred restaurants (Figure 7.4M), indicating changing one's mind was a tougher decision for highly desired rewards. In fact, mice were less capable of choosing to quit before crossing wait-zone thresholds in more-preferred restaurants (Figure 7.4N). This provides a second potential point of decision-making vulnerability in value-conflict between desire and choosing smarter alternatives that could prove problematic in recovering addicts.

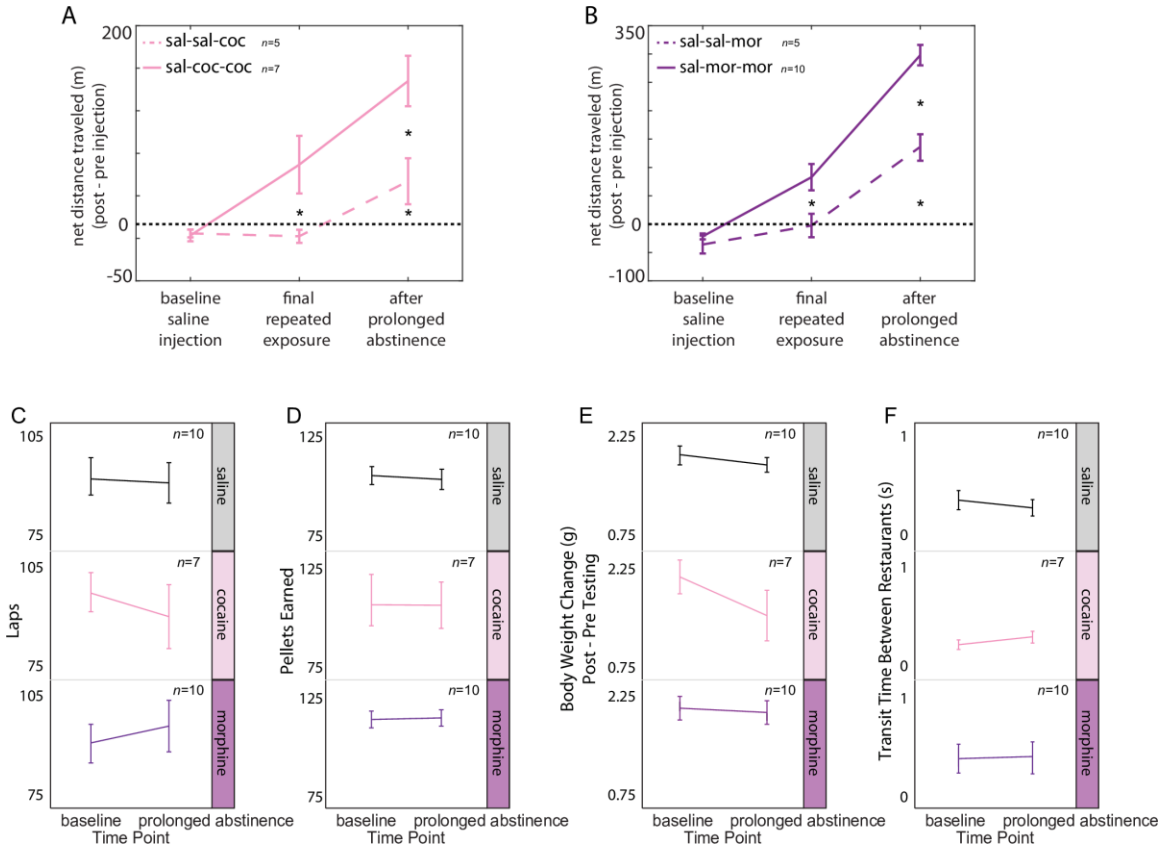
To test potential differences in decision-making vulnerabilities exploited by drugs of abuse, well-trained mice received either repeated cocaine, morphine, or saline using a standard paradigm known to produce psychomotor sensitization (Figure 7.1A, Figure 7.8) – an escalated locomotor response to repeated drug exposure that has been shown to serve as a behavioral correlate of neural plasticity in cortical and mesolimbic pathways, bio-markers of which in humans are predictive of relapse susceptibility (Thomas and Malenka 2003; Thomas et al. 2001; 2000; Ma et al. 2014; Hearing et al. 2018; Wolf 2016; Hearing et al. 2016; Kourrich et al. 2007). Thus, we focused on a timepoint of 2-3 weeks of prolonged abstinence to model the enduring effects of drug use on decision-making processes in recovering addicts.

Importantly, our decision-making tests are made during times when cocaine and morphine are not on board, and we show that drug exposure after the drug has cleared the animal's system does not have any persistent effects on locomotor activity or appetite that could confound our interpretations of our decision-making tests (Figure 7.8).

Acute locomotor and appetite changes are typical effects when drug is on board and could confound behavioral performance on many tasks. The half-life of cocaine is ~1hr and morphine is ~2hr. We tested mice on our task 10 hours after each drug injection (which took place 4 hours post-testing on our task) and well into prolonged abstinence for 2 weeks where we observed our decision-making conflict changes.

We used the following metrics to test for "off-target" effects of chronic drug: speed of locomotion on the task, number of completed laps, total amount of food earned and total weight gained. We found no

Figure 7.8: Psychomotor sensitization and controlling for non-specific drug effects



(A-B) Locomotor activity immediately before and after injections were measured in cocaine- (A) and morphine- (B) treated animals. All animals were injected with saline initially, then mice in the drug groups received repeated respective drug injections while control mice received repeated saline injections (locomotor response to final injection shown). Lastly, following 14 days of prolonged abstinence, enhanced psychomotor sensitization is expressed in mice with a history of drug use, not first-time drug-exposure in saline pre-treated mice. (Friedman, $P<0.05$, post-hoc Mann-Whitney locomotion comparisons between drug groups at time points, $*P<0.05$). (C-F) There were no lasting "off-target" effects on Restaurant Row Performance (C: Laps, D: Pellets Earned, E: Weight Change, F: Transit Speed). All four measures remained constant even after our drug and prolonged abstinence manipulation, implying that these off-target effects did not drive decision-making changes (Friedman, $P>0.05$).

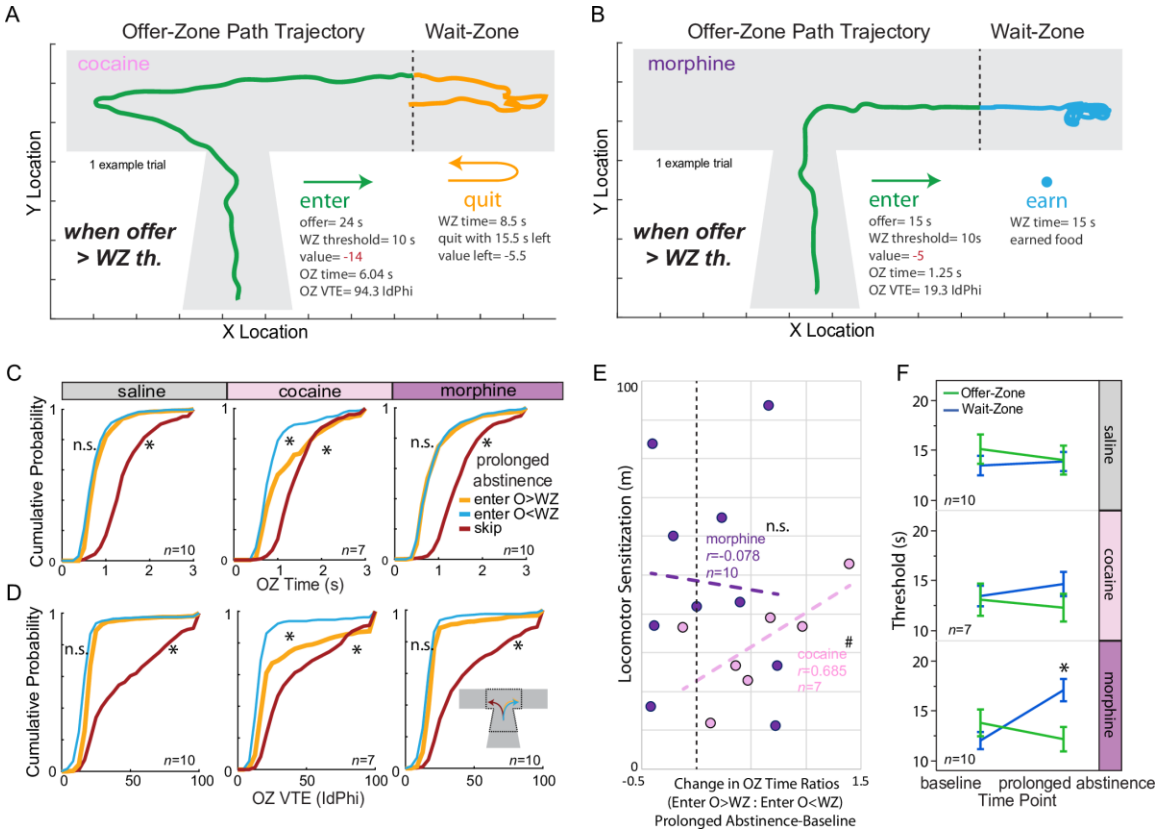
differences in any of these metrics between controls and drug-treated mice (or within individuals) across the entire experiment. This lack of change rules out off-target effects on locomotion or appetite as possible confounding factors for our observed changes in decision-making metrics, including VTE (Figure 7.8).

Interestingly, we found that offer-zone time and VTE were disrupted following prolonged abstinence from repeated cocaine but not morphine or saline exposure (Figure 7.9A,C-D). Cocaine-abstinent mice showed increased deliberation behavior before entering offers greater than wait-zone thresholds, inverting the normal behavior (Figure 7.9A,C-D, compare Figure 7.4I, control simulation analysis Figure 7.10). Cocaine-abstinent mice initially oriented toward skipping these offers, and then re-oriented to accept them anyway (Figure 7.9A). This suggests that cocaine-abstinent mice accepted costly offers despite engaging in VTE and deliberating about turning them down. The degree of change in offer-zone behaviors correlated with degree of psychomotor sensitization in the cocaine group but not the morphine group (Figure 7.9E).

Because entering negatively valued offers ("bad deals" that were typically quit) occurred on a different distribution of offers than entering positively valued offers ("good deals" that were typically earned), we ran simulations that matched the different distributions of offers to each trial type to ensure those differences could not account for changes in offer zone deliberation behaviors that are similar for enters and different for skips at baseline, relationships of which change only in cocaine-abstinent mice (Figure 7.10).

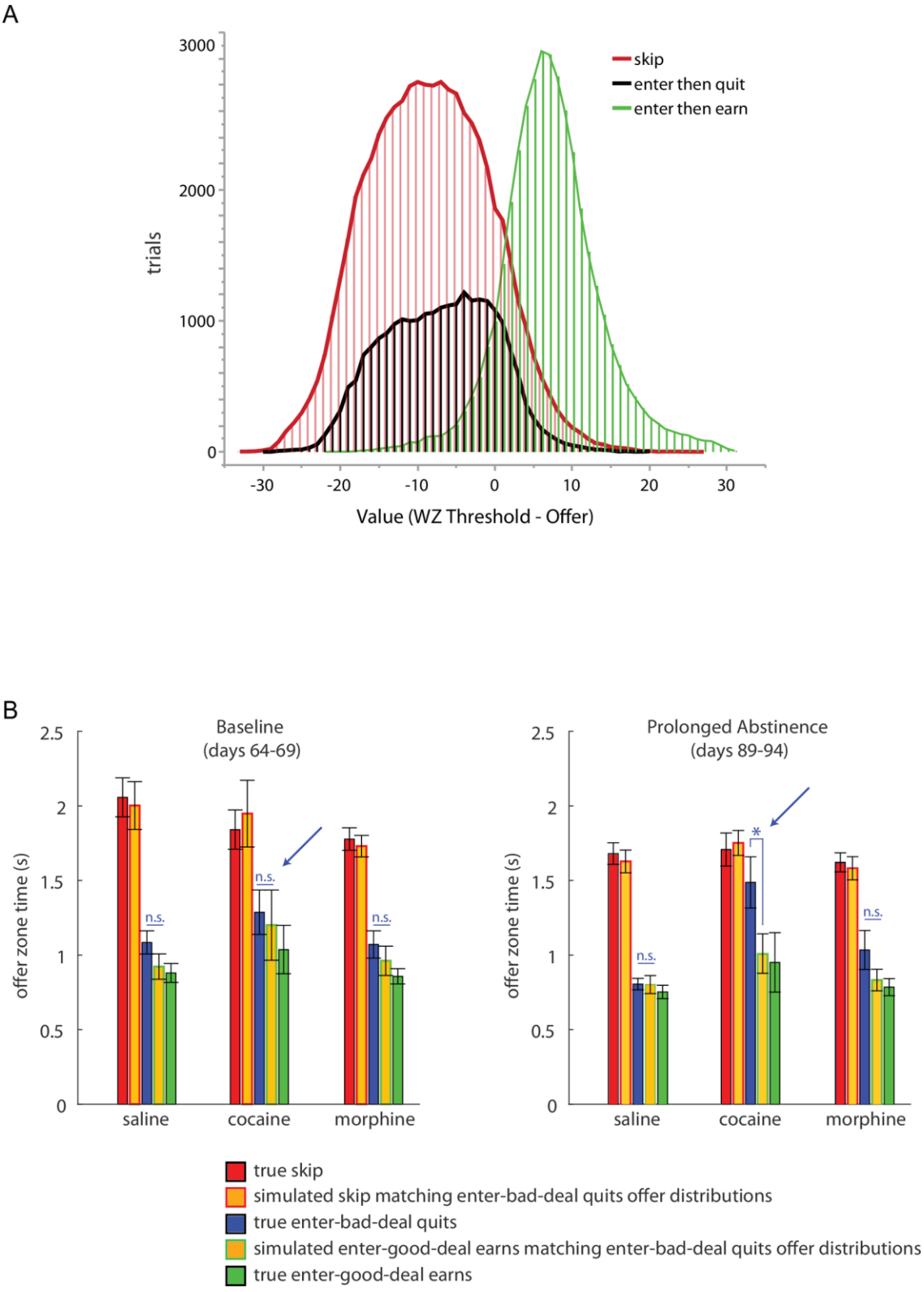
In contrast, morphine-abstinent mice had a significant increase in wait-zone thresholds compared to baseline while cocaine-abstinent and saline-treated mice did not (Figure 7.9F). This is noteworthy because, while morphine-abstinent mice did not differ in making snap-judgments to rapidly accept expensive offers (Figure 7.9C-D), they were less likely to correct those economic violations in the wait-zone in contrast to the saline and cocaine groups (Figure 7.9A-B,F). Thus, probability of quitting significantly decreased (Figure 7.11A). If morphine-abstinent mice did quit, they took significantly longer to do so (Figure 7.11B). Neither cocaine- nor morphine-related effects appeared after a single drug exposure (Figure 7.12).

Figure 7.9: The effects of prolonged abstinence from repeated drug exposure on choice conflict



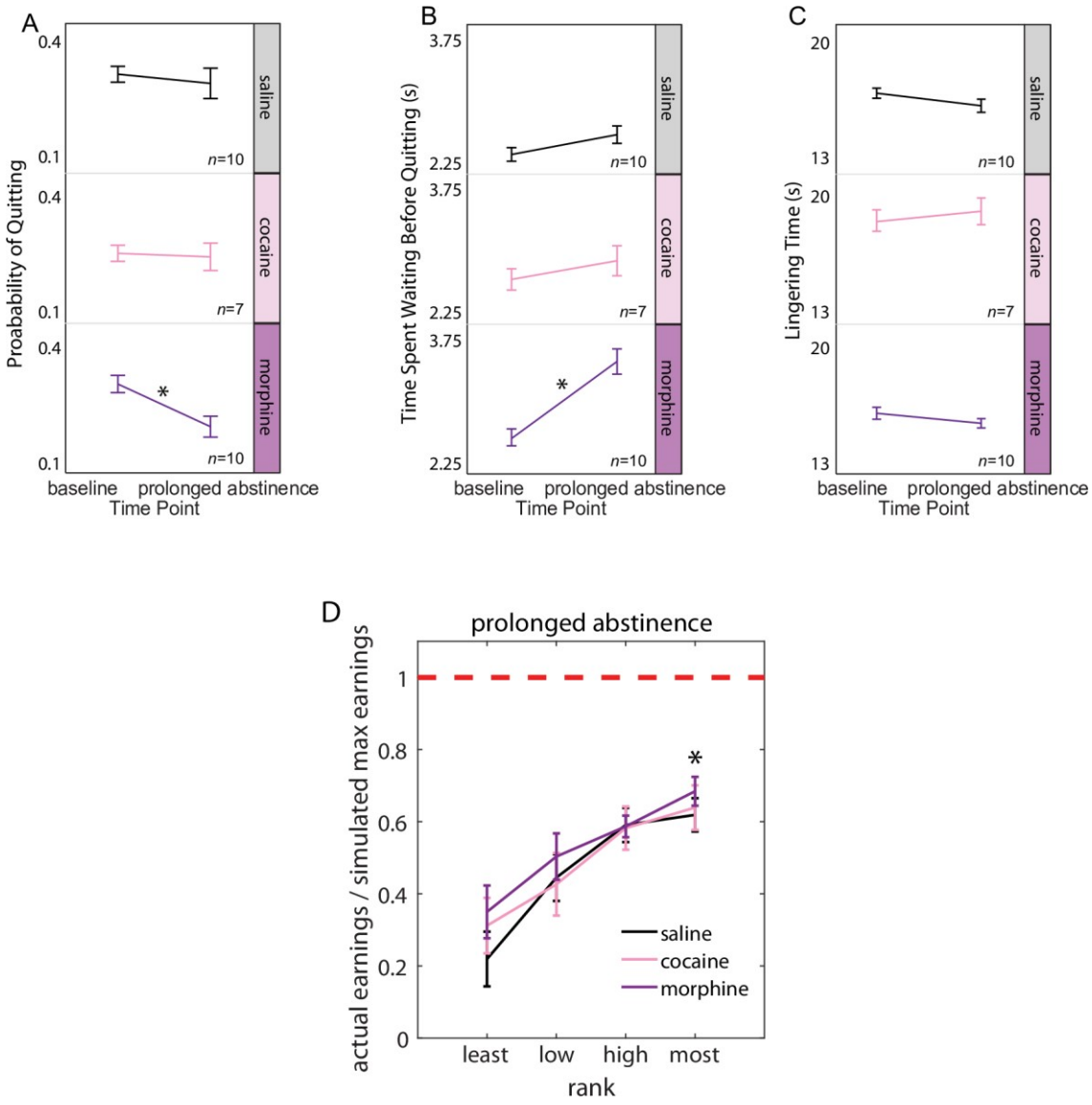
(A-B) Example path-trajectory in the offer- and wait-zones for negatively valued economically unfavorable offers. (A) A mouse with a history of repeated cocaine exposure initially oriented toward skipping (left) then ultimately re-oriented to enter (right). Capability of quitting was unaltered. (B) A mouse with history of repeated morphine exposure was less capable of quitting rapidly accepted offers. (C-D) Cumulative probability distributions of offer-zone time (C) and VTE (D) for skips as well enters split by offer value separated by drug-treatment conditions. Both types of enter decisions were rapid compared to skips and indistinguishable from each other for saline and morphine mice (KS tests, not significant, n.s., P>0.05). Cocaine mice displayed increased time and VTE before accepting negatively valued offers (KS tests, *P<0.05). (E) Relationship between degree of change (from baseline to prolonged abstinence) in the ratio of time to enter negatively valued offers relative to enter positively valued offers correlated with the degree of locomotor sensitization in the cocaine group (Pearson, #P<0.1), not morphine group (Pearson, not significant, n.s., P>0.1). Vertical dashed-line indicates zero change in offer-zone time ratios across experimental timepoints. (F) Repeated measures Friedman tests correcting for multiple post-hoc Mann-Whitney tests revealed wait-zone, not offer-zone, thresholds increased across experimental time points only in the morphine group (*P<0.05).

Figure 7.10: Controlling for offer distribution differences in decision outcomes



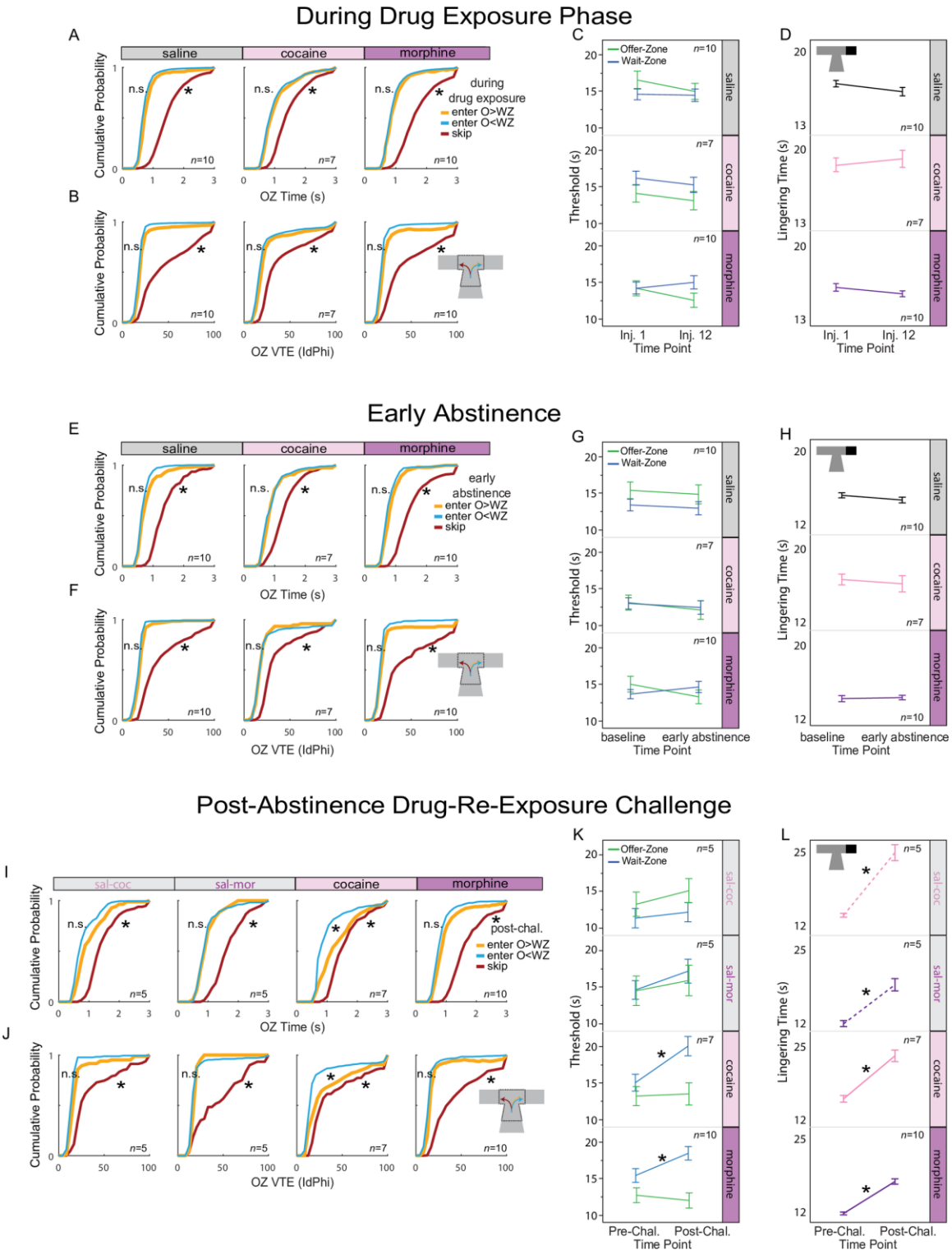
(A) Decision types sorted by trial outcome as a function of offer value (wait zone threshold minus offer) across all animals at baseline intended to illustrate different decision outcomes occur on trials with very different offer distributions, particularly the two types of enter decisions. (B) We ran single trial simulation analyses to control for unequal distributions of offers based on trial type ("skipping bad deal," "entering bad deal then quitting," or "entering good deal") that could confound interpretations of offer zone behaviors when making initial enter or skip decisions. We generated simulated shuffled data sets of both "skipping a bad deal" and "entering a good deal then earning" matching the same trial-by-trial distributions of offer lengths as those subsets of trials where mice "entered a bad deal then quit." That is, simulations were performed by using the offer length distributions that belong to the "enter-bad-deal" scenario and then averaging only those offer-zone reaction times that matched this offer distribution where the outcomes were instead "skips" (for the skip simulation) or "enter-good-deals" (for the enter simulation). We found that after running this analyses on baseline days 64-69, we do not see any significant differences in any treatment group between our conditions of interest (how mice deliberated before accepting bad deals in the offer-zone), comparing "entering bad offers that leads to quits" to the shuffled control "entering that leads to earns simulated to match offer distribution of entering bad deals before quitting," ($P>0.05$). This comparison of interest does change even when matched against simulated shuffled data sets only in the cocaine group after prolonged abstinence ($*P<0.05$). Thus, offer-zone behavior when entering-bad-deals looks like entering-good-deals (both are rapid snap judgments even if the former is a "mistake") for all mice at both time points except the cocaine-treated animals at the prolonged abstinence time point.

Figure 7.11: Effects of prolonged abstinence on additional decision-making metrics



(A) Probability of quitting accepted offers in the wait-zone. (Friedman, morphine *P<0.05, saline/cocaine P>0.05)
(B) Amount of time invested in the wait-zone before quitting. (Friedman, morphine *P<0.05, saline/cocaine P>0.05)
(C) Amount of time spent consuming and lingering at the reward-site post-earning. (Friedman, all groups P>0.05)
(D) Degree of optimal earnings. Sub-optimality was calculated by simulating Restaurant Row sessions and number of max potential earnable pellets using individual thresholds and running speeds, but removing wasteful behaviors (i.e., no quits, no excess time deliberating in offer-zone nor lingering post-earning beyond the minimum showed by the animal). Horizontal dashed red line indicated optimal performance as determined by simulations. (Sign test, P<0.05, all ranks below 1. Kruskal-Wallis-Dunn tests, most-preferred vs. least-preferred *P<0.05).

Figure 7.12: Secondary drug-related timepoints (cyan timepoints 1-3)



Outside of the primary timepoints of comparison in this study (yellow timepoints in Fig.7.1A), we also examined behavior of multiple valuation parameters (offer-zone deliberations, offer- and wait-zone thresholds, and post-earn lingering) during the drug exposure phase (A-D), during early abstinence (E-H), and after the drug-re-exposure challenge intended to assess incubation of psychomotor sensitization (I-L) where mice were either re-exposed to the same drug previously administered or saline-pre-treated mice received either cocaine or morphine for the first time. (A-B) Cumulative probability distributions of offer-zone time (A) and VTE (B) for skips as well enters split by offer value separated by drug-treatment conditions collapsed across the drug exposure sequence. Both types of enter decisions were rapid compared to skip decisions (KS tests, *P<0.05) and indistinguishable from each other (KS tests, not significant, n.s., P>0.05) in all three drug conditions, similar to baseline findings. (C) Friedman tests revealed no changes in offer-zone and wait-zone thresholds across first and last injection of the repeated drug exposure sequence separated by drug-treatment conditions (P>0.05). (D) Similarly, no changes were found in time spent lingering at the reward site after earning across first and last injection of the repeated drug exposure sequence (P>0.05). (E-H) During early abstinence, no changes from baseline were observed in any of the valuation parameters. (E-F) Offer-zone time (E) and VTE (F) for enter decisions were both faster than skips (KS tests, *P<0.05) and indistinguishable from each other (KS tests, not significant, n.s., P>0.05) in all three drug conditions, similar to baseline findings. (G-H) Offer-zone and wait-zone thresholds (G) and lingering behavior (H) did not change over time (Friedman, P>0.05). (I-L) After the drug-re-exposure challenge, although cocaine-treated animals still displayed their main effect (following prolonged abstinence) of increase deliberation time (I) and VTE (J) for offers above wait-zone thresholds, no further changes were seen in any drug condition (KS tests, *P<0.05, not significant, n.s. P>0.05). (K) Only animals with a history of repeated drug exposure (both cocaine and morphine pre-treated groups) displayed increased wait-zone thresholds in response to an acute drug-re-exposure challenge while first-time-exposed mice did not (Friedman, *P<0.05). (L) All mice displayed an increase in lingering behavior following an acute drug challenge (Friedman, *P<0.05).

Our effects of drug on decision-making persist 2 weeks after chronic drug exposure at a time point when long-lasting circuit changes in decision-making-related brain areas including the prefrontal cortex, nucleus accumbens, and hippocampus are known to develop and when psychomotor sensitization is expressed - a hallmark and behavioral correlate of repeated drug-induced incubation of plasticity changes replicated numerous times (Robinson and Berridge 2003; 1993; Hearing et al. 2016; 2018; Wolf 2016; Thomas et al. 2001; Kourrich et al. 2012; 2015; 2007).

Our repeated drug exposure regimen did induce psychomotor sensitization measured in the 90-minute window following drug administration expressed after prolonged abstinence during a drug challenge (Figure 7.8).

Long-lasting changes in decision-making conflict were observed only after repeated drug exposure, not after acute one-time drug exposure (Figure 7.12). We examined behavior during the drug-exposure phase (Figure 7.1A, cyan timepoint 1), during early abstinence (Figure 7.1A, cyan timepoint 2), and following the acute drug-re-exposure change after prolonged abstinence (Figure 7.1A, cyan timepoint 3). The main timepoint of interest was after prolonged abstinence from repeated drug use, a timepoint at which psychomotor sensitization is typically expressed, at which neural plasticity in defined circuits develop, and at which recovering addicts struggle to make good decisions before relapsing (Robinson and Berridge 2003; 1993; Hearing et al. 2016; 2018; Wolf 2016; Thomas et al. 2001; Kourrich et al. 2012; 2015; 2007). Psychomotor sensitization seen after repeated drug exposure has been shown to be a behavioral correlate of drug-induced neural plasticity in specific mesolimbic and striatal circuits. That is, animals that show heightened locomotor responses to drug injections following repeated administration and incubated over prolonged abstinence show drug-induced circuit plasticity while animals that do not show heightened locomotor responses do not exhibit neural plasticity (Rothwell et al. 2011).

Nonetheless, we present additional data during the drug exposure phase, early abstinence, and following the drug-re-exposure challenge primarily intended to express degree of psychomotor sensitization incubated

throughout prolonged abstinence (Figure 7.12). We found no decision-making changes during Restaurant Row during the drug-exposure phase in offer-zone deliberation behaviors (Figure 7.12A-B, enters comparison, non-significant, Kolmogorov-Smirnov tests, $P>0.05$), nor between the first and last (12th) injection during the drug exposure phase in thresholds (Figure 7.12C, wait-zone across time, non-significant, Friedman, $P>0.05$), nor in post-earn lingering time (Figure 7.12D, lingering time across time, non-significant, Friedman, $P>0.05$).

Looking at the early abstinence time point, we found no changes in offer-zone deliberation behaviors (Figure 7.12E-F, enters comparison, non-significant, Kolmogorov-Smirnov tests, $P>0.05$), nor between baseline and early abstinence in thresholds (Figure 7.12G, wait-zone across time, non-significant, Friedman, $P>0.05$), nor in post-earn lingering time (Figure 7.12H, lingering time across time, non-significant, Friedman, $P>0.05$).

Looking immediately following the drug-re-exposure challenge after prolonged abstinence, we only saw the persisting difference in the cocaine group (Figure 7.12I-J, enters comparison, cocaine group only, significant, Kolmogorov-Smirnov tests, $*P<0.05$, see Figure 7.9C-D for comparison). Interestingly, only in mice with a history of repeated drug exposure, and not in formerly saline-treated mice experiencing drug for the first time at the time of the drug challenge, we saw an increase in wait-zone thresholds immediately before and after the drug-re-exposure challenge (Figure 7.12K, wait-zone across time, cocaine and morphine, Friedman, $*P<0.05$). Interestingly, in all mice following the drug challenge, we found an increase in post-earn lingering time (Figure 7.12L, lingering time across time, all mice, Friedman, $*P<0.05$).

Taken together, this suggests that the decision-making changes in mice with a history of repeated cocaine and morphine exposure were apparent only after prolonged abstinence and not after a single drug-exposure. Interestingly, all mice appeared to increase lingering time regardless of history of drug use following an acute exposure to drug (Figure 7.12L). This suggests that hedonic valuations of non-drug rewards can be enhanced during acute withdrawal from drug. An acute drug-re-exposure challenge has been shown in the literature to precipitate reinstatement of drug-seeking behavior as a model of provoking relapse as well as induce neural

plasticity changes unique from prolonged-abstinence-induced plasticity (Robinson and Berridge 2003; 1993; Hearing et al. 2016; 2018; Wolf 2016; Thomas et al. 2001; Kourrich et al. 2012; 2015; 2007). While the main focus of this experiment was not to actually model "triggers of relapse" with drug-re-exposure, but rather model decision-making changes "just before relapse" after prolonged abstinence, it is interesting that drug-re-exposure after prolonged abstinence caused changes in wait-zone thresholds only in mice with a history of repeated drug exposure and not in first-time users (saline-pre-treated mice). This sets the stage for further investigation in future studies to more closely examine the heterogeneity of decision-making changes at secondary timepoint "after" relapse.

Theories of foraging behavior are rooted in hypotheses of optimizing time allocation in order to maximize reward rate (Stephens and Krebs 1986; Charnov 1976). In Restaurant Row, all flavored pellets are of equal caloric value, and thus any differences in reinforcement rate as a function of cost between flavors must be taken as reflecting an underlying subjective valuation. Mice demonstrated a large variability in subjective flavor preferences from which we found interesting asymmetries and interactions with multiple valuation processes measurable on this task.

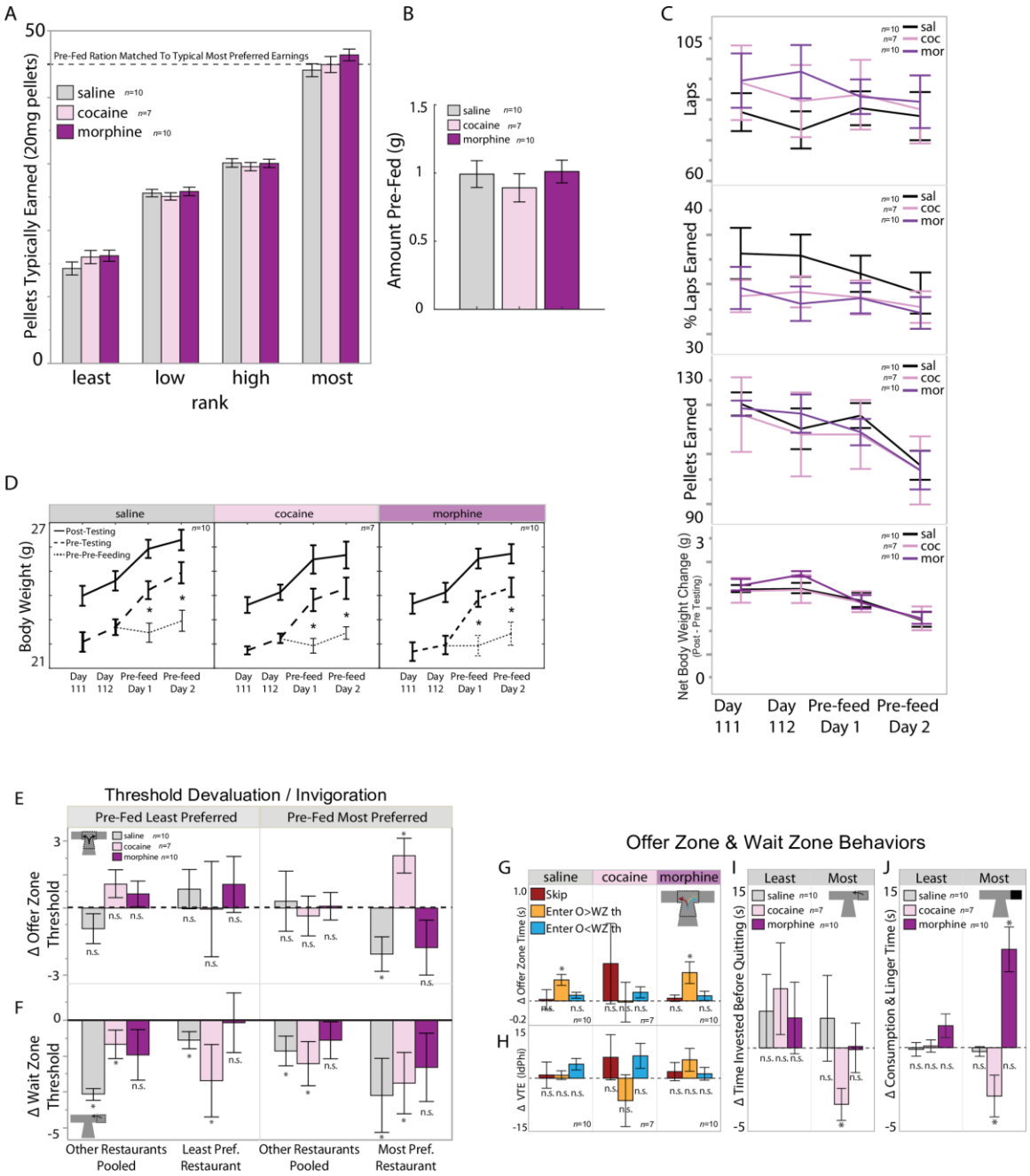
If we take into account individual differences in subjective preferences of willingness to wait for rewards (wait-zone thresholds), we can still determine a measure of sub-optimality, normalized to each animal's idiosyncratic preference for each flavor. In order to calculate maximum number of rewards a mouse could earn in each restaurant taking into account subjective flavor preferences, we simulated Restaurant Row sessions yet eliminated "wasteful" behaviors. To this end, in this model, we forced offer-zone thresholds to match wait-zone thresholds, thus eliminating all quit events. Furthermore, we eliminated differences in offer-zone deliberation time and post-earn lingering time between flavors (by using minimum deliberation time and minimum consumption time collapsed across all restaurants based on each animal's performance). We also used minimum transit time between restaurants based on each animal. These are the times the animal could have used if the only difference between decisions was the underlying willingness-to-wait thresholds between the flavors. We found that mice overall were sub-optimal on this metric, even after taking into

account individual differences in subjective flavor preferences and that prolonged abstinence from repeated drug exposure did not influence this metric (Figure 7.11D).

We also found that degree of sub-optimality interacted with flavor ranking. That is, mice were more sub-optimal in less-preferred restaurants. This is likely due to the disproportionate excess amount of time spent in the offer-zone, wait-zone, and lingering in more-preferred restaurants. Such disproportionate excess amount time that was removed from our optimal-performance model, when re-allocated optimally, would lead the model to predict disproportionately higher earnings than actual in less-preferred restaurants. This is due to the combination of excess time available, lower thresholds in those restaurants, and greater likelihood of our model encountering low cost offers in those restaurants that can be earned and that would have not been actually encountered otherwise. Thus, this yielded higher predicted than actual reinforcement rates in less-preferred restaurants.

Devaluation experiments can modify the incentive value of instrumental actions and reveal specific encoding of emotional states or craving underlying goal-oriented behavior (Colwill and Rescorla 1990; Gremel et al. 2016; Gourley et al. 2016; Wassum et al. 2009a; 2009b; Balleine and Dickinson 1998). In appetitive tasks, pre-feeding is one way to accomplish this. Referring to cyan timepoint 4 in Figure 7.1A and data in Figure 7.13, pre-feeding has been shown to change reward seeking behaviors depending on factors including amount pre-fed, instrumental action being assessed, and reward-selective versus reward-nonselective modulation (Balleine and Dickinson 1998; Wassum et al. 2009b; 2009a; Gourley et al. 2016; Colwill and Rescorla 1990; Gremel et al. 2016). Pre-feeding-induced devaluation of reward-seeking behaviors has been widely used as a way to probe if behaviors are inflexible, stimulus-response-driven, and thus habit-like versus flexible, response-outcome-driven, and thus goal-directed (Balleine and Dickinson 1998; Wassum et al. 2009b; 2009a; Gourley et al. 2016; Colwill and Rescorla 1990; Gremel et al. 2016).

Figure 7.13: Pre-feeding probe session (cyan timepoint 4)



(A-B) Average number of pellets typically earned in each ranked restaurant. Dashed horizontal black line indicates the approximate number of pellets earned in a single session in most-preferred restaurants (A, ~45 pellets, 20mg each). This number was used to determine how much to pre-feed mice before devaluation probe sessions with the intention to partially satiate mice while preserving motivation to run the task following pre-feeding (B, ~0.9g). The same number was also used in both pre-feeding probe sessions regardless if pre-feeding with either the most- or least-preferred flavor. (C) Pre-feeding had no effect on laps run, % laps earned, pellets earned, or net-body-weight-change comparing pre-testing weights to post-testing weights (Friedman, $P>0.05$). (D) Pre-feeding however did increase body-weight when comparing pre-pre-feeding weights to post-pre-feeding weights measured before Restaurant Row testing (Friedman, $*P<0.05$). Pre-feeding plus additional weight gained during Restaurant Row did not significantly change starting pre-pre-feeding weight on the second probe session (Friedman, $P>0.05$). (E-J) We measured changes in behavior of multiple valuation parameters (offer- and wait-zone thresholds, offer-zone deliberations, wait-zone quits, and post-earn lingering) by calculating changes relative to 5d of average behavior preceding the first pre-feeding session (Sign tests, $*P<0.05$, not significant, n.s., $P>0.05$). (F) Mice with a history of repeated saline or cocaine exposure showed decreased wait-zone thresholds in all restaurants in response to pre-feeding regardless of the identity of the pre-fed flavor. Morphine pre-treated mice displayed no changes in wait-zone thresholds. (E) In the offer-zone, thresholds of only the most-preferred flavor only when pre-fed that flavor decreased in saline mice, increased in cocaine mice, and did not change in morphine mice. No other offer-zone thresholds changed in all mice. (G-H) Only saline- and morphine-mice showed increased offer-zone reaction times when accepting offers above wait-zone threshold (G) in the most-preferred restaurant when pre-fed that flavor, however these changes were not accompanied with changes in vicarious trial and error (VTE) behavior (H). (I-J) Cocaine-mice showed a decrease in time invested before quitting (I) and a decrease in time spent lingering (J) in the most-preferred restaurant when pre-fed that flavor while morphine mice only showed an increase in lingering time. Saline mice displayed no changes in wait-zone quit time or post-earn lingering after pre-feeding.

These two potential responses to a devaluation manipulation such as pre-feeding have been shown to separate behaviors that are differentially driven by separable neural circuits.

We pre-fed mice either their least- or most-preferred flavors in an amount that did not disrupt typical number of laps run or pellets earned (Figure 7.13A-C, C: Friedman, non-significant, $P>0.05$). Bodyweight did significantly increase following pre-feeding but before testing, yet was normalized by the next day (Figure 7.13D, before and after feedings, Friedman, significant, $*P<0.05$, before feeding across days, Friedman, non-significant, $P>0.05$).

Wait-zone thresholds were devalued (decreased) in saline and cocaine mice while the thresholds of morphine mice did not change (Figure 7.13F, Sign test, $*P<0.05$). Only when pre-fed their most-preferred flavor were saline mice devalued in the offer-zone as well (Figure 7.13E, Sign test, $*P<0.05$). Offer-zone thresholds of cocaine mice interestingly increased, suggesting pre-feeding for these animals carried an invigorating-like food-prime component on this aspect of behavior (Figure 7.13E, Sign test, $*P<0.05$).

In the offer-zone, deliberation time and VTE when skipping or accepting offers below threshold (economically favorable) was unaltered; however, saline mice accepted offers above threshold (economically unfavorable) more slowly when pre-fed their most-preferred flavor (Figure 7.13G-H, Sign test, $*P<0.05$), suggesting a shift in the balance of valuation functions. Entering offers above threshold however, just as before, took place after little VTE with no further pre-feeding-induced changes, indicating these events were still snap-judgments (did not involve deliberating about correct alternatives, Figure 7.13 H, Sign test, $P>0.05$). Morphine mice responded just as saline mice did while cocaine mice displayed no changes on this metric (Figure 7.13G-H, Sign test, $*P<0.05$).

Finally, although lingering remained unchanged in saline-treated mice, morphine-abstinent mice showed invigorated (increased) lingering while cocaine-abstinent mice showed the opposite (Figure 7.13J, Sign test, $*P<0.05$). Additionally, cocaine-abstinent mice displayed less time spent waiting before quitting (Figure

7.13I, Sign test, *P<0.05). Taken together, pre-feeding revealed changes in dissociable valuation algorithms that were blunted or enhanced based on drug history.

Taking advantage of the subjective value properties of rewards and different zones, we found that pre-feeding decreased wait-zone thresholds (indicating devaluation) consistent with satiety effects on incentive processes. These effects were not-flavor specific and seemed to affect appetitive reward taking valuation processes in general. However, only when pre-feeding most-preferred flavors did offer-zone thresholds also decrease. This highlights not only a flavor-specific satiety effect consistent with past reports but also a subjective value-specific capacity to modify motivational states unique to choose-between decisions involving highly wanted rewards. Pre-feeding seemed to induce invigoration-like effects in drug-treated mice absent in saline-treated mice. In morphine-abstinent mice, we found increased conditioned-place-preference (CPP)-like lingering, which may reflect enhanced craving and explain why their wait-zone thresholds, which were generally insensitive to change, paradoxically opposed satiety-induced devaluation. In contrast, cocaine-abstinent mice, while sensitive to wait-zone threshold devaluation, paradoxically displayed increased offer-zone thresholds. That is, cocaine-abstinent mice were food-primed to over-value offers in the offer-zone that were exaggeratedly under-valued in the wait-zone. Thus, the hypothesis that cocaine-abstinent mice may be transitioning into a lower value state once in the wait-zone may explain why they were more likely to quit, quit faster, and spend less time lingering, suggesting the predicted value of accepted rewards were less than expected.

Pre-feeding was not intended to assess drug effects but rather to assess decision flexibility and rule out habitual processes. Because there were no lasting drug effects on any behavior in the formerly saline animals after the acute drug-re-exposure challenge session which took place 20 days before the pre-feeding probe sessions, this group served as “control” conditions for the pre-feeding probe. Again, the fact that all groups still showed sensitivity to the pre-feeding probe (although with intricate fine-grained differences between groups), we determined that the decision-processes in Restaurant Row remained flexible and had not transitioned to habit-like processes.

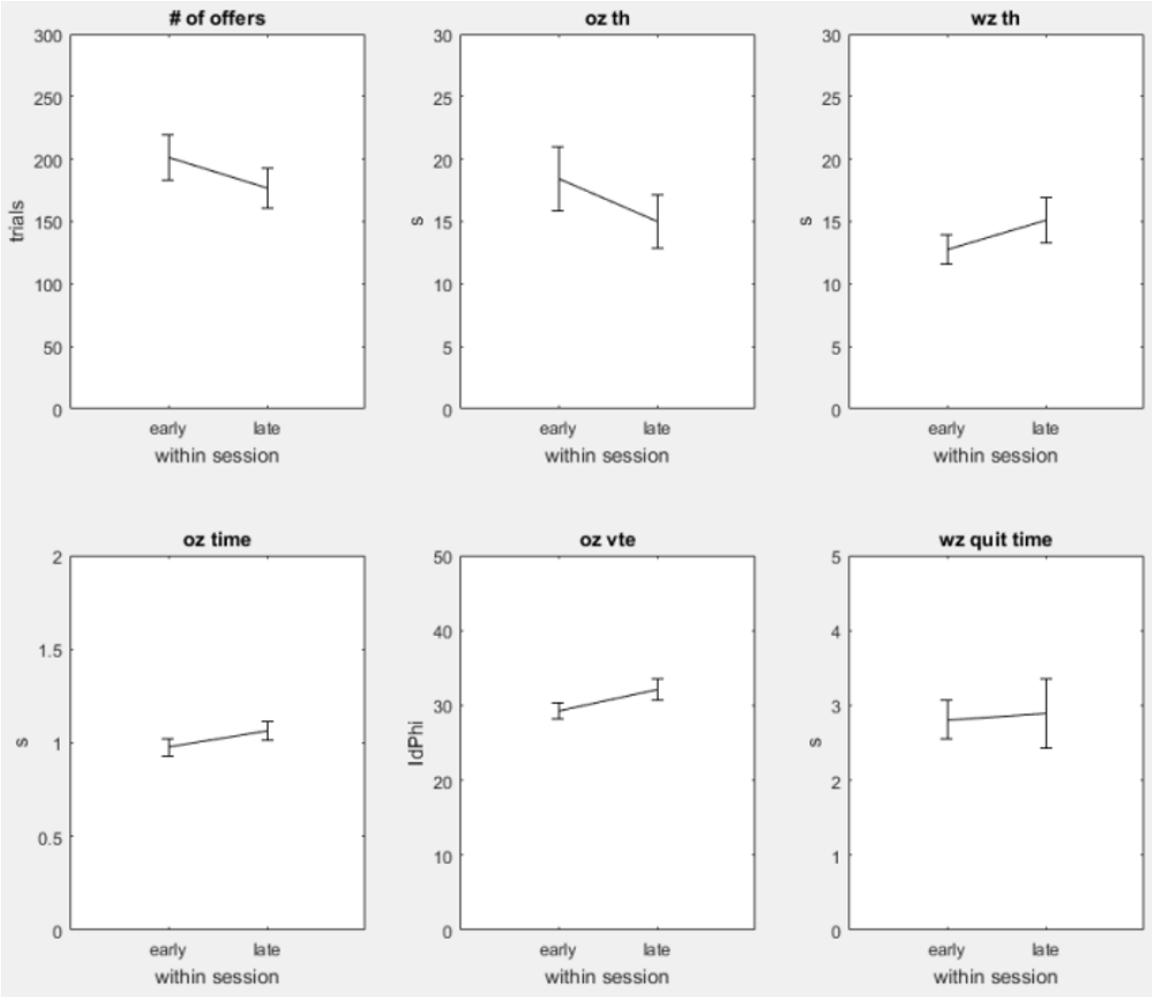
Related to satiety, we checked to see if within session feeding imposed satiety like effects from session start to session end. To do this, we compared a number of behavioral metrics during the first 30 min to the last 30 min of a session. We analyzed behavior from well-trained mice on day 70. There were no dynamic changes in behavior within session, for example, due to within-session fatigue or within-session satiety (not due to pre-feeding devaluation but regular Restaurant Row (Figure 7.14).

Discussion

Recent findings have suggested that choosing between distant options accesses different valuation processes than choosing to opt out from remaining committed to already accepted offers (Carter and Redish 2016). We can model such decision framings as fundamentally distinct types of intertemporal choice modalities.

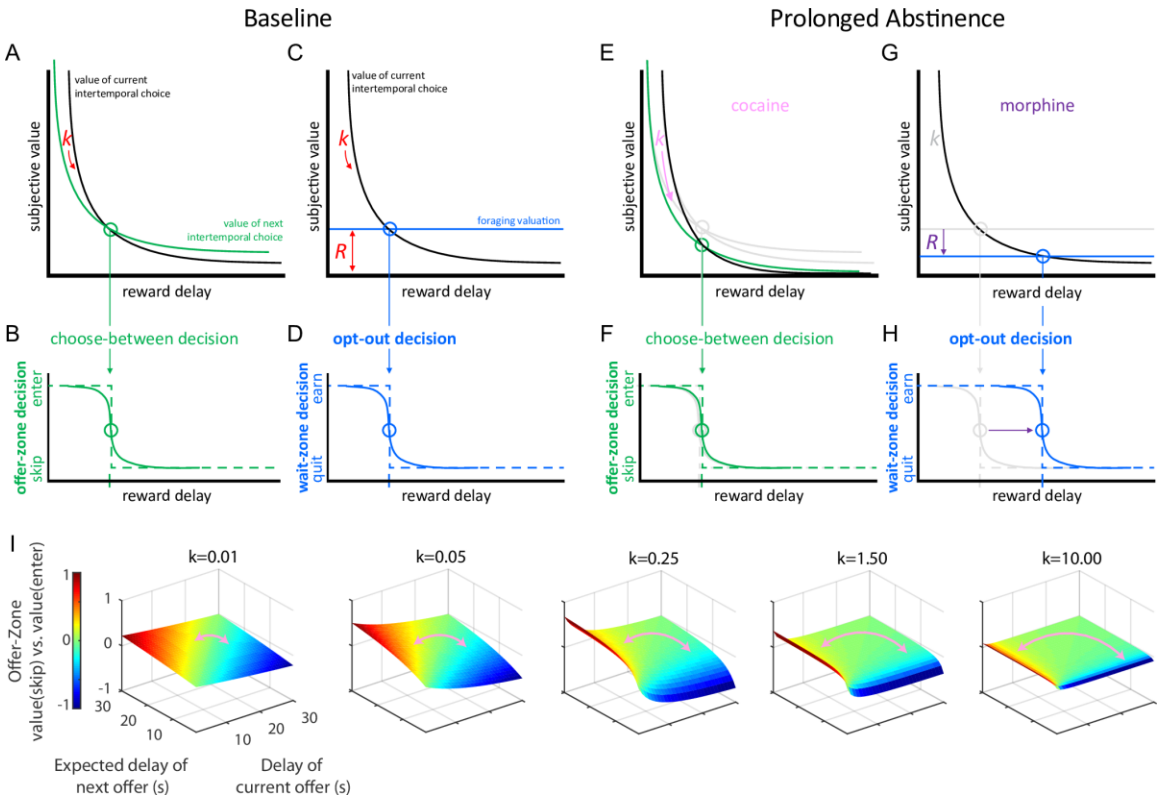
Because VTE behavior occurs in the offer-zone, animals are likely to be engaged in episodic future thinking and deliberation (Redish 2016). During VTE, hippocampal representations sweep forward along the path of the animal, alternating between potential goals (Johnson and Redish 2007). Such goal representations are synchronized to reward value representations in the prefrontal cortex and ventral striatum, suggesting outcome predictions are being evaluated serially during VTE (van der Meer and Redish 2009; Steiner and Redish 2012). This is dissociable from dorsal striatum valuations that occur during rapid decisions when VTE is not engaged (van der Meer et al. 2010). To this end, we modeled two hyperbolic functions discounting the value of the known current and expected next alternative. The decision change occurs at the intersection of these two hyperbolic functions (Figure 7.15A). This well-established neuroeconomic model of choosing between alternatives underlies the offer-zone threshold valuation measured on our task (Figure 7.15B, Ainslie 1975; Glimcher et al. 2007; Kable 2007).

Figure 7.14: Lack of an effect of within-trial time on Restaurant Row behaviors



We split the 1 hour session into the first 30 min (early) and last 30 min (late) and assessed if there were any changes on 6 different metrics (number of offers encountered, offer zone thresholds, wait zone thresholds, offer zone decision time, offer zone vicarious trial and error [VTE], and time spent in the wait zone before quitting). There are no changes within session across any of these metrics, showing that there are no effects of satiety within session.

Figure 7.15: Neuroeconomic modeling of separable valuation algorithms



(A-D) Baseline. (A) Deliberative model: hyperbolic temporal discounting function of the current choice (black) is compared against a second hyperbolic temporally discounted function of the expected next choice (green), with a discounting rate "k" (red). (B) Offer-zone "choose-between" thresholds are derived from this intersection. (C) Foraging model: hyperbolic temporal discounting function (black) of "work remaining" with discounting rate "k" (red) is compared against the average opportunity cost of reward availability in the rest of the environment, y-intercept "R" (red). (D) Wait-zone "opt-out" thresholds are derived from this intersection. (E-I) Modeling the effect of our drug delivery and forced abstinence manipulation. (E) Our data in mice with a history of repeated cocaine exposure is consistent with an increase in the "k" parameter in offer-zone deliberative valuation model, which yields no change in offer-zone thresholds (F), but yields increased indecision particularly for economically unfavorable high cost offers (I). (F) Our data in mice with a history of repeated morphine exposure is consistent with a decrease in the "R" parameter in the wait-zone foraging valuation model, which leads to an increase in the wait-zone threshold (H).

In contrast, quitting the wait-zone is an "opt-out" decision. Such "is it worth it?" judgments appear in well-studied decision-processes common in foraging paradigms (Carter and Redish 2016; Wikenheiser et al. 2013; Stephens and Krebs 1986; Charnov 1976). This can be modeled as a comparison of the hyperbolic temporally discounted value of "work remaining" compared against the average opportunity cost of reward availability in the rest of the environment (Figure 7.15C). The intersection of this comparison underlies the wait-zone threshold valuation measured on our task (Figure 7.15D).

Studies have modeled changes in the hyperbolic discounting rate "k" in drug users as steeper, thus over-valuing immediate rewards (Bickel et al. 2015). These tasks, however, do not typically characterize the deliberation behaviors that led up to the outcomes selected. Other theories have proposed that the average available reward "R" undergoes a normalization step (decreases) in addiction thus decreasing the value of alternative options in the rest of the environment (Moal and Koob 2007). Importantly, both of these valuation changes (an increase in "k" or a decrease in "R") could drive recovering addicts to make bad decisions and relapse (Redish et al. 2008).

Our data revealed no changes in either the offer-zone or wait-zone threshold in cocaine-abstinent animals. From this, we must conclude that whatever decision-making changes occurred in the cocaine-abstinent animals, it did not shift the cross-over points in deliberative or foraging valuation algorithms. What we did find is an increase in offer-zone deliberations for costly offers. This effect could occur as a consequence of a change (increase) in offer-zone "choose-between" hyperbolic discounting rate "k" (Figure 7.15E-F,I). Because hyperbolic discounting curves decrease in steepness as one moves out along the curve, this would effectively decrease discriminatory resolution when "choosing-between" costly offers (Figure 7.15I).

Our data revealed no change in the offer-zone threshold, but did find a right-shift in the wait-zone threshold of morphine-abstinent animals. This cannot occur due to an increase in the hyperbolic discounting rate "k" because such a change would decrease the wait-zone threshold (see foraging model, Figure 7.15C-D). Instead, this right-shift in the willingness to wait out a delay once started is exactly what we would expect to see if the effect of morphine was to diminish the average rate of reward "R" expected in the world (Figure

7.15G-H), consistent with recent theories of opiate abuse (Redish et al. 2008). Taken together, I highlight two dissociable points of failure in decision-making exploited uniquely by two drugs of abuse – before making "bad" deliberative judgments versus re-evaluations after making "bad" snap judgments.

These findings are particularly relevant to a timepoint when recovering addicts who are on the verge of relapse struggle with making the right decisions. Our work highlights the notion that complex valuation processes can be carefully modeled in animal behavior. Disruptions in deliberative processes separate from foraging processes can suggest distinct circuit-specific computations that can go awry in different forms of addiction.

Many studies examining the lasting neurobiological changes induced by different drugs of abuse, including psychostimulants and opiates, generally propose a unified theory of addiction common to most abused substances that converges on overlapping changes in synaptic plasticity within the mesolimbic reward system (Hearing et al 2018). The majority of these studies focus on changes in glutamatergic and dopaminergic signaling in the ventral tegmental area and nucleus accumbens (Hearing et al 2018). However, there are reports of contrasting or opposing lasting neurobiological changes induced by cocaine and morphine, including differential effects on accumbens spine density, synaptic remodeling, and gene expression (Alcantara et al 2011; Russo et al 2011; Robinson and Kolb 2004; Becker et al 2017). I suggest that taking into account the information processed within these circuits as well as other circuits during discrete aspects of decision-making computations is critical in order to understand multi-faceted, potentially dysfunctional valuation processes that can ultimately drive addiction-related behaviors.

This data uncovers unique computation-specific etiologies separated within the same trial that may be underlying different forms of addiction that more traditional behavioral paradigms may not be sensitive enough to detect. I propose that computation-specific therapeutic interventions are likely necessary to ameliorate addiction sub-types that disrupt, in different ways, the decision to use despite knowing better.

– Interlogue –

Chapter 8

Resolving disease heterogeneity via disruption in distinct valuation algorithms

In the previous chapter, I demonstrated how prolonged abstinence from chronic exposure to cocaine or morphine produced long-lasting dissociable changes in distinct aspects of decision-making information processing in mice tested in this novel variant of the Restaurant Row task.

First, this task allows us to operationalize a sophisticated level of decision-conflict – the conflict between "wanting" vs. "knowing better" – measurable through behavior on this neuroeconomic task. This task models complex decisions that recovering human addicts struggle with before relapsing. However, these processes are poorly understood and have not been well-modeled in animal decision making in previous tasks.

Second, I segregated stages of this conflict into distinct deliberative valuation algorithms (in the offer zone) separate from foraging valuation algorithms (in the wait zone). In chapters two and four, I took a neuroeconomic approach to demonstrated how the information represented in a deliberative algorithm is fundamentally distinct from the information represented in a foraging algorithm, and that this distinction can be separated behaviorally. In chapter six, I discussed a new theoretical framework combining a neuroeconomic approach with a disease models that move beyond simple tests of value in order to reveal more about neuropsychiatric disease heterogeneity in dissociable computational processes. In the previous chapter, I directly tested this idea comparing how two different drugs of abuse could give rise to fundamentally distinct changes in separable decision processes. I discovered that prolonged abstinence from

chronic exposure to drugs of abuse disrupt these types of decision conflict at a time point relevant to when recovering addicts struggle with relapse.

Cocaine-abstinent mice displayed disruptions in offer zone behaviors before giving in to wanting despite knowing better during high-conflict scenarios. This could be modeled as a disruption in a deliberative algorithm. Morphine-abstinent mice, after making initial snap-judgement mistakes in the offer zone, displayed distinct disruptions in wait zone behaviors during high-conflict scenarios. Morphine-abstinent mice were less likely to make corrective change-of-mind decisions, a process that remained intact in cocaine-abstinent mice. This could be modeled as a unique disruption in a distinct foraging algorithm. Importantly, these lasting behavioral changes took place a time point when lasting changes in neural plasticity have been reported.

In numerous animal models of addiction, prolonged abstinence from drugs of abuse can give rise to lasting changes in neural plasticity. Such plasticity changes. regardless of the drug of abuse tested, are often reported to be similar in nature. For instance, long-term potentiation in corticostriatal synapses develops following abstinence from chronic drug exposure. However, our results presented in the previous chapter is at odds with a dogma generally accepted in the field of addiction research – that different forms of addiction all converge on similar pathology and maladaptations in behavior. Here, I found that abstinence from cocaine vs. abstinence from morphine separately disrupted fundamentally distinct decision-making valuation algorithms. These changes were specific to decision conflicts in deliberative vs. foraging algorithms. This suggests that although there may be overlapping circuit changes induced by different types of drugs of abuse, distinct failure modes in decision making could likely to be underlying fundamentally distinct pathologies in different types of addiction.

Taken together, these data reveal that heterogeneity in addiction-related changes in decision-making information processing can be resolved using neuroeconomic paradigms that move beyond simple tests of value.

In the next chapter, I will test a critical commentary made in chapter 6 – direct circuit-specific manipulations of plasticity delivered acutely in an "off-line" fashion, that is, delivered outside of behavioral testing in a neuroeconomic paradigm in order to capture circuit-specific neural computations in distinct behaviors separated across space and time.

– Original Research –

Chapter 9

Altering gain of the infralimbic to accumbens shell circuit alters economically dissociable decision-making algorithms

Abstract

The nucleus accumbens shell (NAcSh) is involved in reward valuation. Excitatory projections from infralimbic cortex (IL) to NAcSh undergo synaptic remodeling in rodent models of addiction and enable the extinction of disadvantageous behaviors. However, how the strength of synaptic transmission of the IL-NAcSh circuit affects decision-making information processing and reward valuation remains unknown, particularly because these processes can conflict within a given trial, and particularly given recent data suggesting that decisions arise from separable information-processing algorithms. The approach of many neuromodulation studies is to disrupt information flow during on-going behaviors, however this limits interpretation of endogenous encoding of computational processes. Furthermore, many studies are limited by the use of simple behavioral tests of value, which are unable to dissociate neurally distinct decision-making algorithms. I optogenetically altered the strength of synaptic transmission between glutamatergic IL-NAcSh projections in mice trained on a novel neuroeconomic task capable of separating multiple valuation processes. I found that induction of long-term depression (LTD) in these synapses produced lasting changes in foraging processes without disrupting deliberative processes. Mice displayed inflated re-evaluations to stay when deciding whether or not to abandon continued reward-seeking investments but displayed no changes during initial commitment decisions. I also developed a novel ensemble-level measure of circuit-specific plasticity that revealed individual differences in foraging valuation tendencies. Our results demonstrate that alterations in projection-specific synaptic strength between IL-NAcSh is capable of augmenting self-control economic valuations within a particular decision-making modality and suggests that the valuation mechanisms for these multiple decision-making modalities arise from different circuits.

Chapter reprinted with permissions from *PNAS*, modified from:

Sweis BM, Larson EB, Redish AD, Thomas MJ. 2018e Altering gain of the infralimbic to accumbens shell circuit alters economically dissociable decision-making algorithms. *PNAS* 201803084.

Introduction

Addiction, like many other neuropsychiatric disorders, is often characterized by an inability to regulate maladaptive motivated behaviors, particularly after continued drug use (Robinson and Berridge 2003). Recent advancements in neuroscience have begun to dissect apart reward-related circuitry in order to understand how specific circuits change over time during addiction (Morales and Margolis 2017; Lüscher and Malenka 2011). Targeting such disorders with plasticity-altering manipulations could prove extremely useful to provide long-lasting treatments that can modify disease trajectory and prevent recovering addicts from making poor decisions that lead to relapse.

There is an intimate link between memory and decision making (Redish and Mizumori 2015). Weights of synaptic inputs change as a consequence of learning and carry information, including experiences related to addiction (Hyman 2005; Hyman and Malenka 2001). How information is stored changes how it is processed during decision-making (Redish and Mizumori 2015). However, understanding how the synaptic efficacy of a specific circuit impacts distinct aspects of decision-making information processing has been largely unexplored. Only recently have tools been developed to directly manipulate the plasticity of specific inputs. Thus, direct interrogation of the synaptic efficacy of specific circuits is critical for our understanding of decision-making information processing and disease etiology. Such approaches will be critical for the development of lasting disease-augmenting neuromodulation therapies.

The shell of the nucleus accumbens (NAcSh) is a specialized region of the striatum where reward valuations are encoded and where motivation is thought to be translated into reward-seeking actions (van der Meer et al. 2010; Stott and Redish 2014; German and Fields 2007). The NAcSh receives dense, direct excitatory input projections from the infralimbic (IL) cortex, a sub-region of the prefrontal cortex that has been implicated in the learning and maintenance of cognitive flexibility and self-control (Barker et al. 2014). Thus, the IL-NAcSh circuit is in a key position to serve as a critical conduit mediating information flow that integrates regulatory higher-order cognitive processes with reward-seeking motivated behavioral output processes (Barker et al. 2014).

The IL-NAcSh circuit is highly susceptible to the effects of drug-related experiences. In humans, functional connectivity of an analogous circuit is augmented in recovering drug users (Camchong et al. 2014). Functional magnetic resonance imaging studies that followed abstinent drug users found that strength of functional connectivity (coherence in resting-state BOLD signal) between the medial prefrontal cortex and nucleus accumbens was weaker in individuals that relapsed sooner, consistent with a hypothesized prefrontal role in overriding accumbens-driven relapse behaviors (Camchong et al. 2014). In parallel, several animal models of addiction produce plasticity in the synaptic efficacy in the IL to NAcSh circuit, particularly at times of decision-making vulnerabilities that often lead to drug relapse (Hearing et al 2016, 2018, Creed et al 2015, Pascoli et al 2014, Pascoli et al 2011, Thomas et al 2001). These data suggest that a key factor in addiction lies in experience-dependent circuit plasticity across this circuit, both over the course of abstinence from drugs of abuse as individuals transition into long-lasting decision-vulnerable states as well as immediately following an acute trigger of relapse. Thus, it is clear that the IL-NAcSh circuit has an important role in addiction etiology. However, current theories of decision-making suggest that decision arises from multiple separable decision-making systems, and current theories of addiction suggest that addiction can arise from multiple failure points within those decision-making systems (Redish 2013; Redish et al. 2008; Walters and Redish 2018; Rangel et al. 2008; Bickel et al. 2012). How the IL-NAcSh circuit interacts with different decision-points remains unknown.

Current approaches in systems neuroscience have not directly interrogated the functional consequences of circuit-specific synaptic remodeling on decision-making information processing. Instead, the majority of recent neuromodulation studies manipulate brain activity during on-going behaviors. Direct optogenetic inhibition of the IL-NAcSh pathway can cause spontaneous reinstatement or enhance cue-induced reinstatement of extinguished reward-seeking behaviors while excitation of the IL-NAcSh pathway during cue presentation can block reinstatement of extinguished reward-seeking behaviors (Peters et al. 2008; Gutman et al. 2017; Barker et al. 2014; Augur et al. 2016; Gass and Chandler 2013; Peters et al. 2009). Such studies can easily probe temporally precise necessity and sufficiency of the IL-NAcSh circuit in discrete

regulatory processes. However, such "on-line" manipulations during on-going behaviors in real time impose disruptions of endogenous neural signaling and provide little insight into the functional consequences of synaptic remodeling on behavior.

An alternative approach is to alter the synaptic efficacy of signal transmission specifically of the IL-NAcSh circuit through optogenetically-driven alterations in synaptic plasticity. Importantly, the goal of this approach is to change the gain of the information endogenously transmitted through this circuit, not disrupt the information itself that is coded in this specific circuit. Thus, these types of manipulations should be delivered "off-line" outside of behavioral testing. Such an approach leaves endogenous information processing intact during behavioral observations measured at a later timepoint, when the functional consequences of lasting changes in circuit gain are realized. To date, only a handful of studies have directly augmented strength of synaptic transmission of glutamatergic IL afferents in the NAcSh (Pascoli et al. 2011; 2014; Creed et al. 2015; Hearing et al. 2016). However, these studies have revealed conflicting findings, suppressing reward-seeking reinstatement in some cases while precipitating reinstatement in others.

Importantly, these studies have relied on relatively simple behavioral tests of value and compulsive drug-seeking behavior. Recent theories in neuroeconomics suggest that decisions made in different situations derive from different valuation functions residing in separable neural circuits (Redish 2013; Rangel et al. 2008). It can be difficult to behaviorally segregate these parallel information processing algorithms using traditional behavioral tasks (Kalenscher and van Wingerden 2011). Neurally distinct computations can produce what appears to be (superficially) similar behaviors. Unless a task is specifically designed to separate them, behavioral consequences of distinct neural computations can appear grossly similar and thus remain unseparable. In order to discriminate neural computations through behavior and thus reveal circuit-specific information processing, I applied "off-line" plasticity manipulations of the IL-NAcSh circuit in mice who learned a complex neuroeconomic task that is capable of separating behavioral consequences among different decision-making systems.

Here, I combine these two approaches: circuit-specific "off-line" manipulations of strength of IL-NAcSh synaptic transmission with complex behavioral testing designed to dissociate existing neuroeconomic theories of parallel valuation systems. I adopted a novel neuroeconomic task, Restaurant Row, for use in mice that separates deliberative valuation algorithms from foraging valuation algorithms for natural rewards within the same trial (Sweis et al. 2018b; 2018c Steiner and Redish 2014, Chapters 2 & 4). I observed changes in separable aspects of behavior in foraging valuations but not deliberative valuations that may more closely reflect changes in dissociable neural computations that underlie those dissociable behaviors. In doing so, I also discovered and developed a straightforward way to measure individual differences in projection-specific strength of synaptic transmission at the neuronal population ensemble level.

Methods

Mice

30 C57BL/J6 male mice (wild type), 13 weeks of age, were trained in Restaurant Row. 16 additional behavior-naïve mice were used solely for electrophysiology tests. Mice were single-housed (to protect skull implants) in a temperature- and humidity-controlled environment with a 12-hr-light/12-hr-dark cycle with water ad libitum unless being testing in Restaurant Row in which case only water was ad libitum and food (full nutrition flavored pellets) was task-earned. Mice were food restricted to a maximum of 85% free feeding body weight and trained to earn their entire day's food ration during their 1-hr Restaurant Row testing session. All experiments were approved by the University of Minnesota Institutional Animal Care and Use Committee. Research technicians were blinded to animal conditions. Previous studied using this task yielded reliable behavioral findings with minimal variability in at least sample sizes of n=5 mice per experimental group. Electrophysiology findings from our lab report reliable data from at least n=6 slices.

Surgery

Animals were anesthetized with a ketamine and xylazine mixture (100 and 10 mg/kg; i.p.) and placed into the stereotaxic frame (Kopf Instruments). Virus AAV8-Syn-Chronos-GFP (University of North Carolina Vector Core Facility) was bilaterally transfected into infralimbic cortex (IL, stereotaxic coordinates adapted

from mouse brain atlas: anterior-posterior: +1.65; medial-lateral: ±0.4; dorsal-ventral:−3.2 from bregma) of 7-8 wk old mice. Injections of virus (0.5 μL per injection site) were performed with a 5-μL Hamilton syringe using an UltraMicroPump with SYS-Micro4 controller (World Precision Instruments). After a 5-min delay to reduce solution backflow along the infusion track, the syringe was slowly removed over a 5-min period and incisions were closed using Vetbond (3M). Approximately 1 wk following virus surgery, mice were bilaterally implanted with indwelling optic fibers (200/230 μm core/cladding, 0.66 NA) targeting the accumbens shell (NAcSh, angled: 14 degrees; anterior-posterior: +1.5 mm; medial-lateral: ±1.63; dorsal-ventral: −4.1) using stereotaxic apparatus described above. Light guides were secured to the skull with a dual-cure resin-lonomer (Geristore) that was anchored with two 0.0625-in steel machine screws (amazon.com) with caps (ThorLabs) placed over the top. After approximately 4 wk of recovery, mice began training in behavior while a separate group was yoked and remained housed without behavior training for an equal amount of time before all electrophysiology experiments were conducted in all mice at the same time to match expression level profiles.

Electrophysiology

Parasagittal slices (240 μm) containing the NAcSh were sliced in an ice-cold solution saturated with 95% O2/5% CO2 containing 75mM sucrose, 87mM NaCl, 2.5mM KCl, 1.25mM NaH2PO4, 26mM NaHCO3, 3mM Na ascorbate, 7mM MgSO4, 0.5mM CaCl2. Slices recovered for at least 60 min in a room-temperature artificial CSF solution saturated with 95% O2/5% CO2 containing 119mM NaCl, 2.5mM KCl, 1.0mM NaH2PO4, 1.3mM MgSO4, 2.5mM CaCl2, 26.2mM NaHCO3, and 11mM glucose. For all electrophysiological recordings, picrotoxin (Sigma-Aldrich, 100μM) was added to block GABAergic neurotransmission. Using an Axon Instruments Multiclamp 700B (Molecular Devices), extracellular field recordings were sampled through a pulled glass pipette electrode filled with aCSF. Recording electrode was placed in NAcSh. Data were filtered at 2 kHz and digitized at 5 kHz via custom Igor Pro software (Wavemetrics). Blue LEDs (Plexon, 465 nm, 15 mW) were used to drive optical light pulses, delivered into the slice preparation via a 1 meter patch cable (Plexon) connected to a polished bare fiber tip lowered into the field of view and a focal spotlight pulse directed over the tissue surrounding the recording electrode.

Pulses were delivered via a Master-8 stimulator (A.M.P.I.). Optically evoked field potentials of the IL-NAcSh circuit were obtained every 15s. Pulse durations ranged from 0.1-4ms. Peak amplitude of the N1 and N2 components of the waveform were calculated from average traces. Pharmacology tests were run by washing in DNQX (Sigma-Aldrich, 1μM, 10μM) or TTX (Tocris, 1μM). %Change in N1 and N2 were measured relative to baseline averages before drugs were washed in or calcium levels changed in the zero-calcium test. Plasticity tests were run by stimulating at a constant stimulus duration determined to elicit a 50% max response for 12.5 min (50 pulses) before delivering either no stimulation (10min), 10 Hz stimulation (10min, 4ms pulses), or 100 Hz stimulation (1s train, 4ms pulses, 4 trains, 10s inter-train-interval) protocols. Following one of these three protocols, regular sampling at the original constant stimulus duration resumed for 37.5 min (150 pulses). %Change in N1 and N2 were measured in the final 50 pulses compared relative to the initial 50 baseline pulses. Synaptic strength assays were carried out by ramping pulses up from 0.1 ms to 1.0 ms in 0.1 ms increments (10 steps) collecting 5 pulses at each step, and then ramping back down, in total collecting 100 pulses taking 25 min. Peak N1 and N2 amplitude pairs were calculated from each stimulus duration step and scatter-plotted against each other. Slope from linear regression was calculated as a metric of strength of synaptic transmission. Slopes were collected before and after plasticity induction protocols to measure change in strength of synaptic transmission as a consequence of slice exposure to no stimulation, 10 Hz, or 100 Hz. Additional mice received either no stimulation, 10 Hz, or 100 Hz in vivo and then were sacrificed 24 hr later for functional connectivity assays to be compared across animals between groups.

Behavior

Mice underwent 1 week of pellet training prior to the start of being introduced to the Restaurant Row maze. During this period, mice were taken off of regular rodent chow and introduced to a single daily serving of BioServ full nutrition 20 mg dustless precision pellets in excess (5g). This serving consisted of a mixture of chocolate-, banana-, grape-, and plain-flavored pellets. Next, mice (hungry, before being fed their daily ration) were introduced to the Restaurant Row maze 1 day prior to the start of training and were allowed to roam freely for 15 min to explore, get comfortable with the maze, and familiarize themselves with the feeding

sites. Restaurants were marked with unique spatial cues. Feeding bowls in each restaurant were filled with excess food on this introduction day.

Task training was broken into 4 stages. Each daily session lasted for 1 hr. At test start, one restaurant was randomly selected to be the starting restaurant where an offer was made if mice entered that restaurant's T-shaped offer zone from the appropriate direction in a counter-clockwise manner. During the first stage (day 1-7), mice were trained for 1 week being given only 1 s offers. Brief low pitch tones (4000 Hz, 500 ms) sounded upon entry into the offer zone and repeated every second until mice skipped or until mice entered the wait zone after which a pellet was dispensed. To discourage mice from leaving earned pellets uneaten, motorized feeding bowls cleared any uneaten pellets upon restaurant exit. Left over pellets were counted after each session and mice quickly learned to not leave the reward site without consuming earned pellets. The next restaurant in the counter-clockwise sequence was always and only the next available restaurant where an offer could be made such that mice learned to run laps encountering offers across all four restaurants in a fixed order serially in a single lap. During the second stage (day 8-12), mice were given offers that ranged from 1 s to 5 s (4000 Hz to 5548 Hz, in 387 Hz steps) for 5 days. Offers were pseudo-randomly selected such that all 5 offer lengths were encountered in 5 consecutive trials before being re-shuffled, selected independently between restaurants. Again, offer tones repeated every second in the offer zone indefinitely until either a skip or enter decision was made. In this stage and subsequent stages, in the wait zone, 500ms tones descended in pitch every second by 387 Hz steps counting down to pellet delivery. If the wait zone was exited at any point during the countdown, the tone ceased and the trial ended, forcing mice to proceed to the next restaurant. Stage 3 (day 13-17) consisted of offers from 1 s to 15 s (4000 Hz to 9418 Hz) for another 5 days. Stage 4 (day 18-70) offers ranged from 1s to 30s (4000 Hz to 15223 Hz) and lasted until mice showed stable economic behaviors. We used 4 Audiotek tweeters positioned next to each restaurant powered by Lepy amplifiers to play local tones at 70dB in each restaurant. We recorded speaker quality to verify frequency playback fidelity. We used Med Associates 20mg feeder pellet dispensers and 3D-printed feeding bowl receptacles fashioned with mini-servos to control automated clearance of uneaten pellets. Animal tracking, task programming, and maze operation was powered by AnyMaze (Stoelting). Mice were tested at the same

time every day in a dim-lit room, were weighed before and after every testing session, and were fed a small post-session ration in a separate waiting chamber on rare occasions as needed to prevent extremely low weights according to IACUC standards (not <85% free-feeding weights).

After 70 consecutive days of testing in Restaurant Row, mice were divided into 3 groups of 10 and received either no stimulation (10 min), 10 Hz stimulation (10 min, 4 ms pulses), or 100 Hz stimulation (1 s train, 4 ms pulses, 4 trains, 10 s inter-train-interval) protocols in the evening 4 hours after Restaurant Row testing completed on day 71 for round 1 of stimulations. Mice were transported to a separate stimulation room and plugged into fiber optic patch cables (Plexon, 1 meter) bilaterally connected to compact magnetic LC blue LEDs (Plexon, 465 nm, 15 mW) on a commutator swivel. Mice resumed regular daily testing in Restaurant Row. On the evening of day 81, mice received round 2 of stimulations, where the 10 Hz and 100 Hz groups swapped stimulation protocols in a cross-over repeated measures design. On the evening of day 93, mice received round 3 of stimulations according to original protocol assignments for 5 consecutive evenings. Mice were tested up until day 106 and retired before being sacrificed for electrophysiological assessments.

Statistical analysis

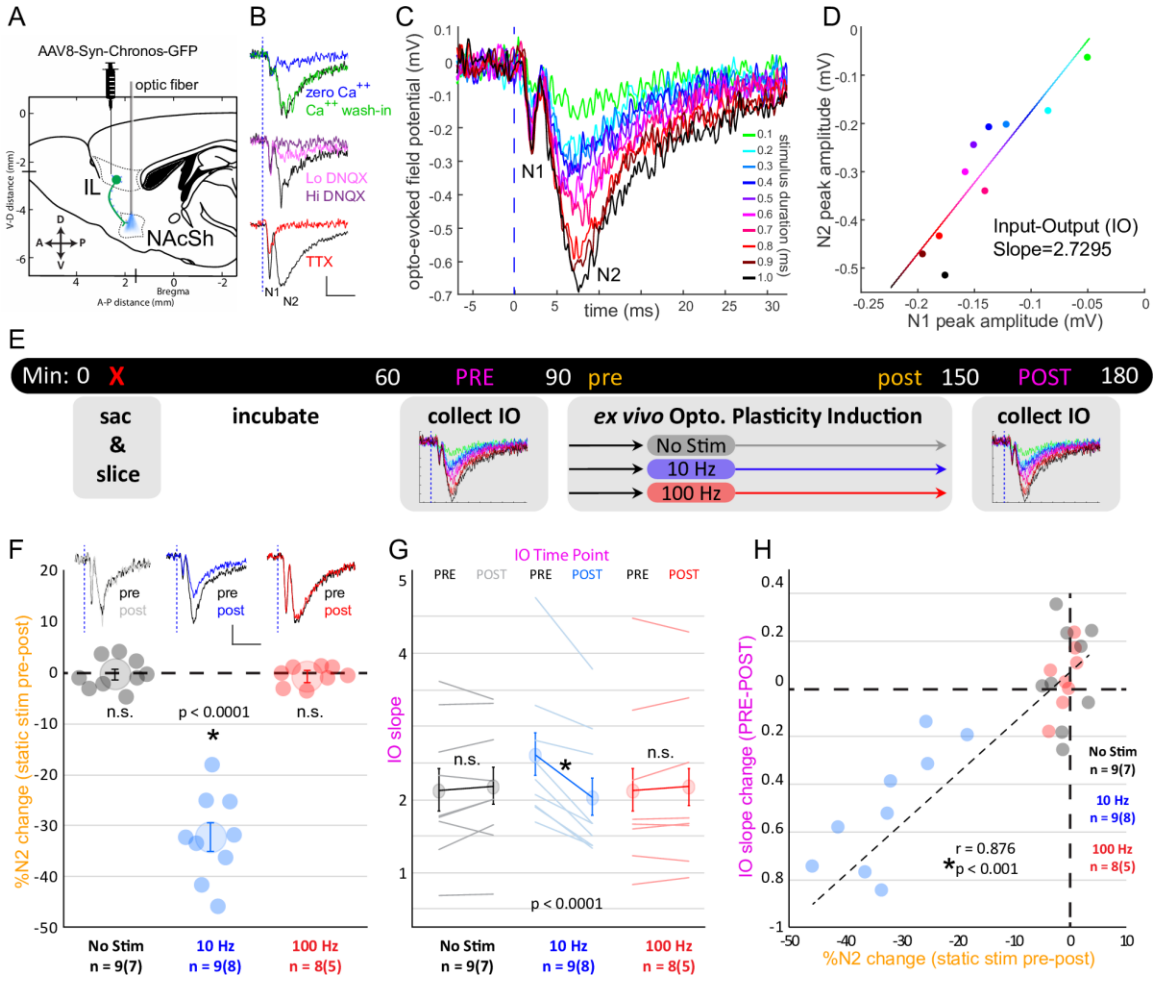
All data were processed in Matlab and statistical analyses were carried out using JMP Pro 13 Statistical Discovery software package from SAS. All data are expressed as mean +/- 1 standard error. Sample size is included in each figure, where electrophysiology data are reported n=slices(mice). Statistical significance was assessed using student's t tests, one-way, two-way, and repeated measures ANOVAs, with post-hoc Tukey t tests correcting for multiple comparisons. Correlations were reported using Pearson correlation r coefficients. Electrophysiology waveforms represent plotted averages from 10-50 consecutive pulsed evoked responses, described in detail in each corresponding figure.

Results

To gain input-specific control of the IL-NAcSh circuit, we bilaterally transfected IL neurons in male mice via intracranial infusions of an adeno-associated viral (AAV8) construct containing the gene for Chronos (Klapoetke et al. 2014), a light-gated fast-kinetics cation-channel opsin, fused to a green fluorescent reporter protein driven by the neuron-specific synapsin (Syn) promoter (Figure 9.1A, Figure 9.2). This promoter directed Chronos expression specifically to neurons, projections of which out of IL are largely comprised of excitatory, pyramidal glutamatergic efferent axons. Then, we bilaterally implanted mice with optic fiber ferrules directed at NAcSh intended to illuminate and activate Chronos-containing axon terminals originating from IL with blue light *in vivo*. Furthermore, IL projections to the ventral striatum preferentially innervate NAcSh while nearby cortical regions like the prelimbic cortex preferentially innervate the nucleus accumbens core (Britt et al. 2012; Brog et al. 1993). This difference in anatomical topographical organization helped ensure that unintentional virus transfection outside of IL was less likely to cross-contaminate manipulation of the target IL-NAcSh circuit. We then allowed mice to recover to allow for robust expression and anterograde transport of Chronos downstream along IL axons delivering opsins to IL terminals in NAcSh.

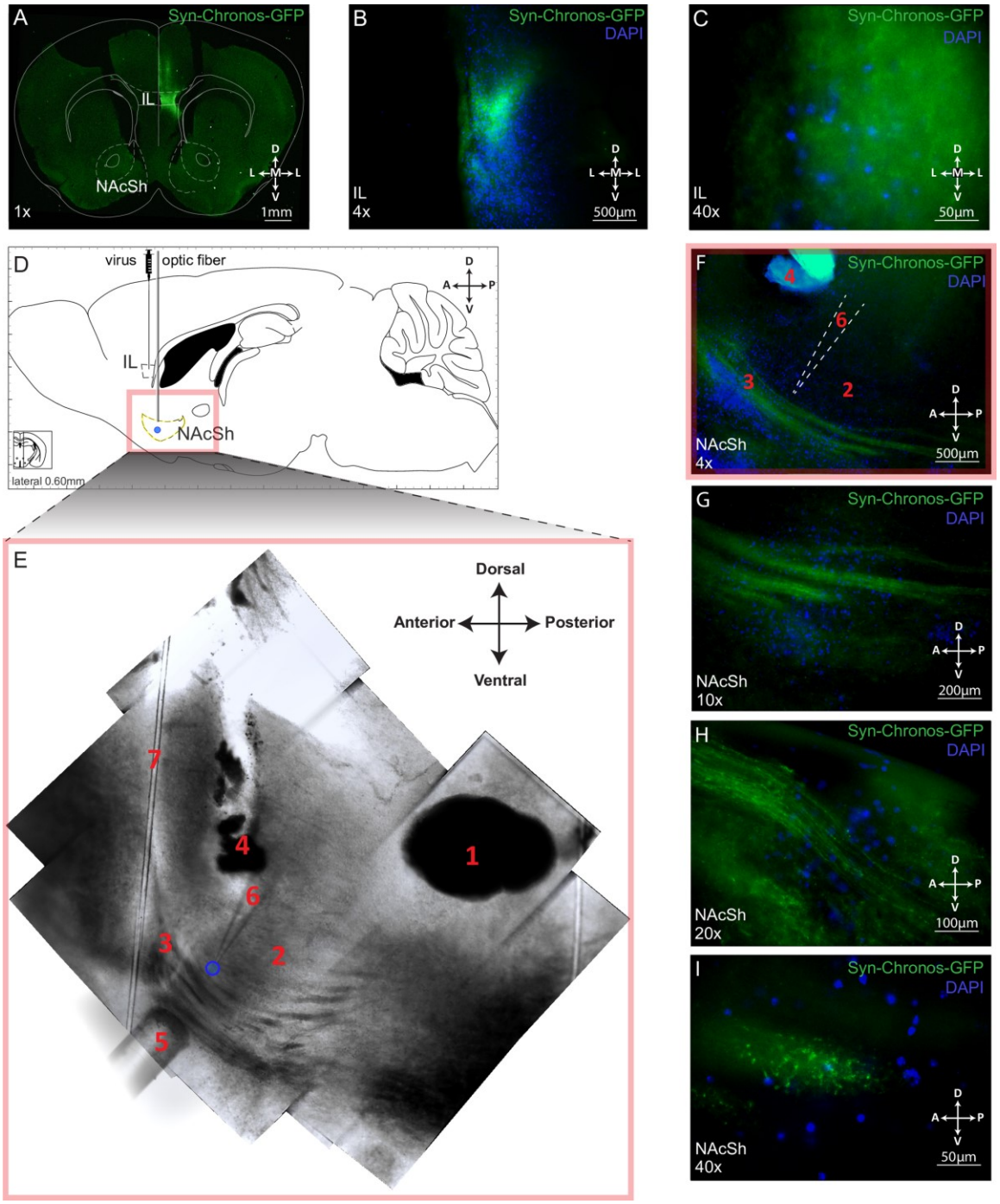
To test the effect of our IL-NAcSh circuit manipulation, we directly recorded electrophysiology of the IL-NAcSh circuit from parasagittal brain slices that contained NAcSh and preserved functional IL axon projections (Figure 9.1B, Figure 9.3). The purposes of performing these experiments in slice *ex vivo* was to ensure we were directly recording in NAcSh and could visualize dense IL afferent fiber bundles. Additionally, this approach also allowed us to test concurrent, serial pharmacological assays that together characterize the waveform of optogenetic-evoked field potentials in this circuit not previously demonstrated. By placing an extracellular recording electrode directly in NAcSh and illuminating a focal area of the surrounding tissue with blue light, we recorded light-evoked population potentials of the IL-NAcSh circuit. This reliably evoked a population potential with a waveform consisting of two negative peaks, labeled N1 and N2 (Figure 9.1B, Figure 9.4). We found that only the N2 component required

Figure 9.1: Optogenetic induction of circuit-specific plasticity in IL-NAcSh at the ensemble level



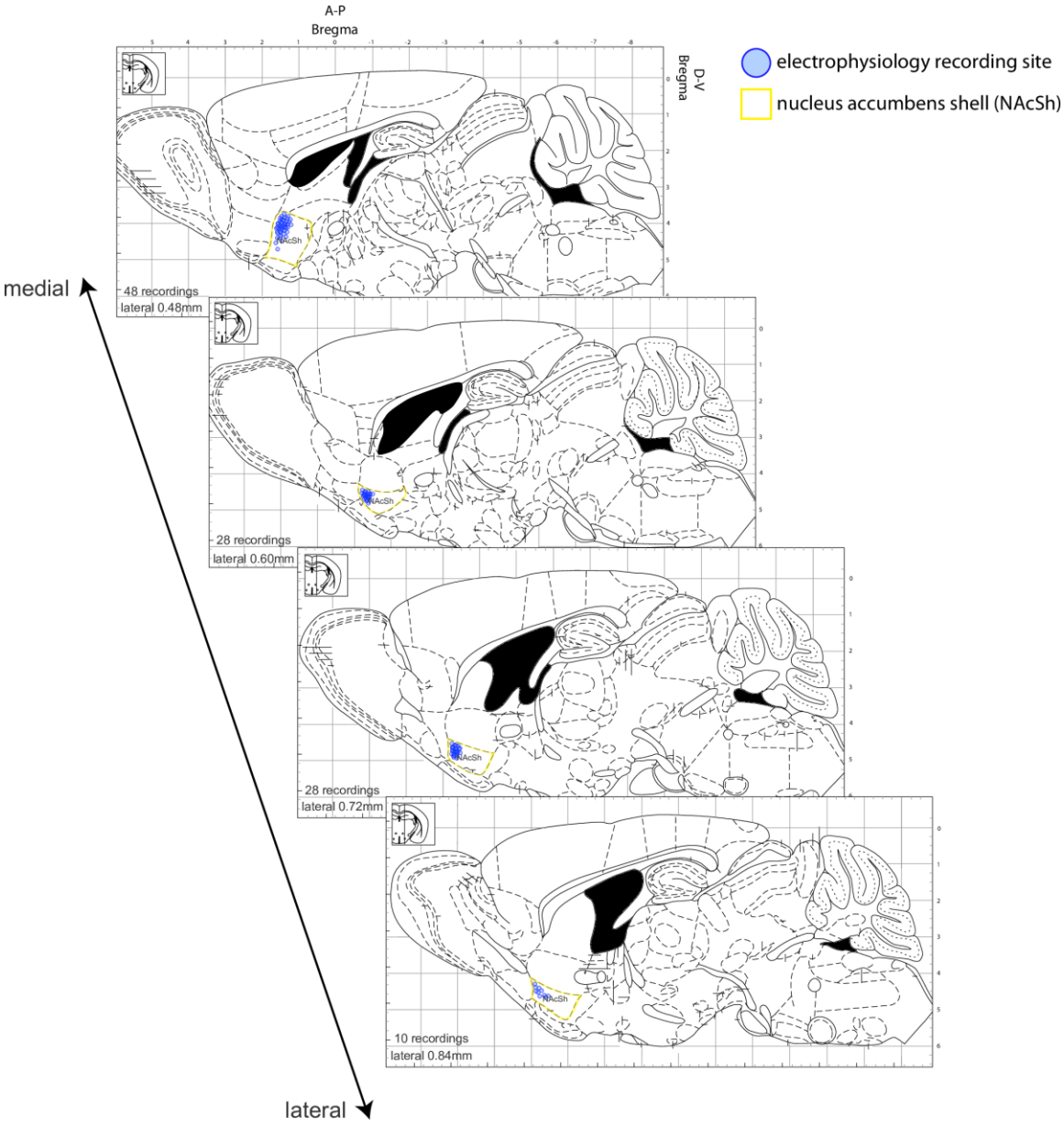
(A) Surgical schematic, parasagittal view. Mice were bilaterally transfected in IL with an adeno-associated virus (AAV8) construct containing the gene for Chronos driven by a neuron-specific synapsin (Syn) promoter with a fusion green fluorescent reporter protein (GFP). (B) Pharmacological characterization of bimodal light-evoked field potentials recorded ex vivo in NAcSh. Vertical dashed blue line indicates 1ms light pulse. Top: zero calcium bath reversibly abolished N2 component. Middle: AMPA receptor antagonist abolished the N2 component, dose-dependently. Bottom: TTX abolished the N2 while reducing the N1 component. Scale: x=10ms, y=0.2mV. (C) Example input-output (IO) regimen assayed on a single slice. Stimulus duration was varied from 0.1 ms to 1.0 ms. (D) N1 and N2 peak amplitudes are plotted from example traces in (C) and linear regression slope was calculated. To test if IO slope can serve as an independent metric of circuit-specific synaptic strength, IO ramps were assayed before and after ex vivo bath application of plasticity-inducing optogenetic stimulation protocols. (E) Experimental timeline. Mice were sacrificed and slices were prepared at time zero. After allowing 1hr for tissue incubation, the first IO slope (PRE) was collected. From this, the stimulus duration that elicited 50% max N2 amplitude was determined and used as the static stimulus parameter during regular samplings for the bath-application plasticity assay (12.5min baseline sampling [pre] followed by one of three protocols and then 37.5min post-protocol sampling. Final 12.5min served as [post]). A second, post-plasticity IO slope assay was collected (POST). (F) Percent N2 change normalized to baseline sampling [pre-post]. Representative wave form traces plotted from pre and post time points for each protocol. Scale: x=10ms, y=0.2mV. (G) IO slopes from PRE and POST time points. (H) %N2 change in (F, pre-post) plotted against change in IO slope in (G, PRE-POST) with Pearson correlation, r. Dots in (F,H) and thin lines in (G) represent individual slices. Sample size is noted near the x-axis, as number of slices followed by number of mice in parentheses. Larger dots represent mean +/- 1 standard error. *p value on plot, indicates not significant (n.s.).

Figure 9.2: Projection-specific targeting of infralimbic to accumbens shell circuit



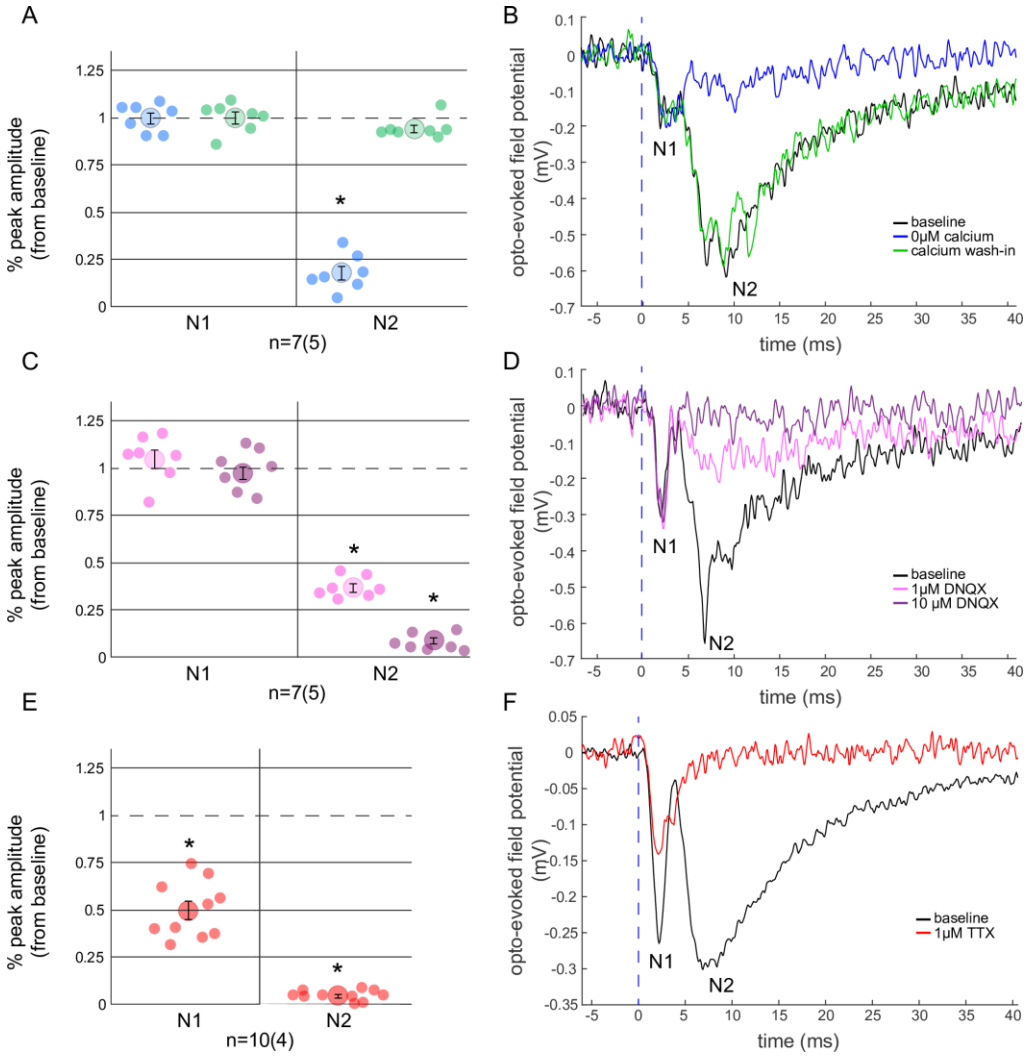
(A) Coronal section of demonstration mouse unilaterally transfected in IL in the right hemisphere only (mice used in the experiment were bilaterally transfected). IL at the virus transfection site, at 4x (C) and 40x (D). IL cell bodies stained with DAPI. (D) Parasagittal schematic of viral transfection site in IL and optic fiber implantation site in NAcSh (outlined in yellow). Blue dot indicates location of ex vivo recording site. (E) Red box zoom in from (D) depicts light-field microscope image collage from a live recording session in NAcSh under light-field microscope. Representative landmarks and instruments annotated by 1-7. (1) Anterior commissure reference landmark. (2) NAcSh region. (3) High density fiber bundle at the anterior border of the NAcSh. GFP label in fluorescent images in (G) reveal these to be virus-transfected opsin-containing IL afferents. (4) Remnant optic fiber track from in vivo implantation. (5) Working optic fiber tip lowered into the slice for ex vivo illumination. (6) Extra-cellular recording electrode pipette lowered into the slice ex vivo. (7) Cloth harp strand used to hold the slice in place while in the bath. (F) Red box zoom in from (D-E) depicts NAcSh at 4x. Number annotations follow that in (E). Estimated recording pipette location outlined in white. NAcSh cell bodies stained with DAPI. NAcSh imaged at 10x (G), 20x(H), and 40x(I).

Figure 9.3: Localization of IL-NAcSh electrophysiology recording measurements



Parasagittal sections taken from the Paxinos mouse brain atlas, with NAcSh outlined in yellow in each section. Translucent blue circles represent recording site locations.

Figure 9.4: Characterization of waveform components of optogenetic evoked IL-NAcSh field potentials

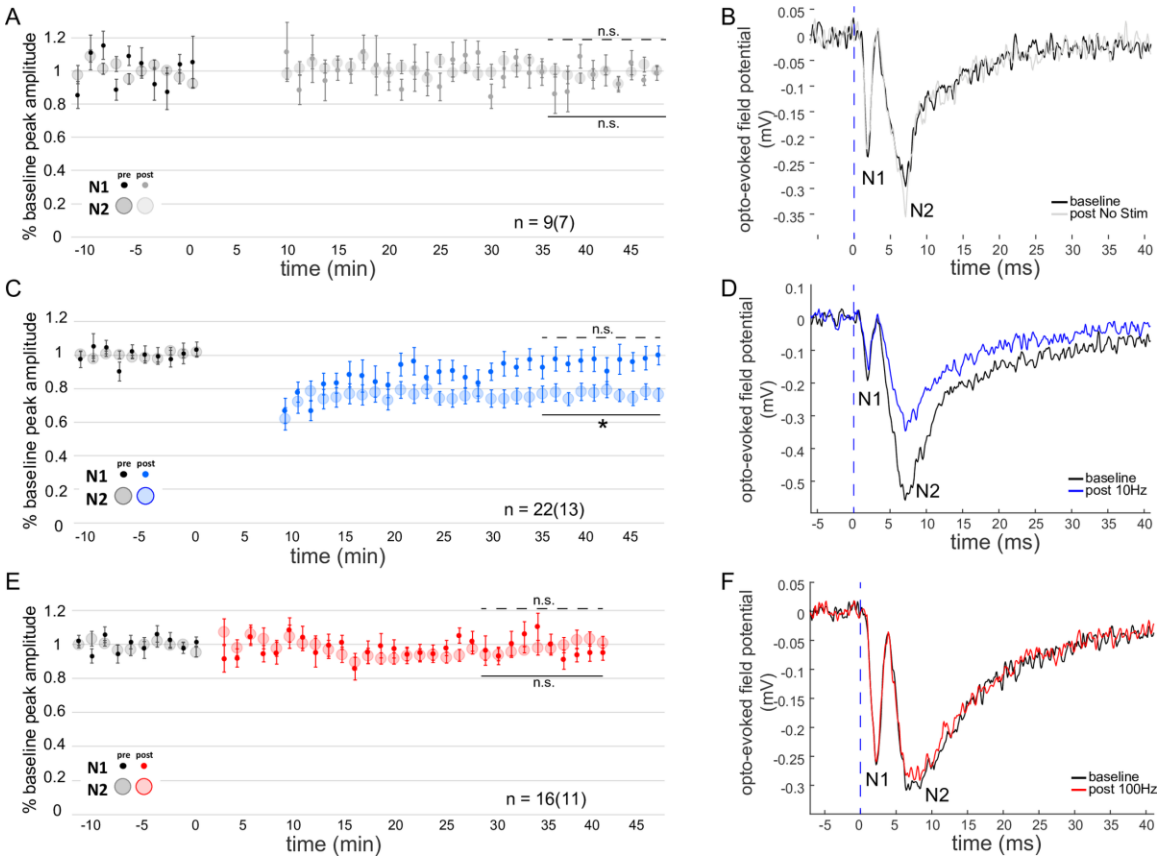


Brief light pulse (1 ms) evoked a bimodal field potential waveform without eliciting a stimulus artifact recorded extracellularly in NAcSh slices. First and second negative peaks were labeled N1 and N2 respectively. Vertical dashed blue line represents onset of light pulse at time zero. (A-B) Switch to a 0μM calcium bath reduced only the peak N2 amplitude component without affecting the N1 component. N2 component was rescued following calcium wash-in. (C-D). Application of increasing bath concentrations of DNQX (AMPA receptor antagonist) reduced only the N2 component in a dose-dependent manner without affecting the N1 component. (E-F) Application of TTX (voltage-gated sodium channel antagonist) abolished the N2 component while reducing the N1 component. Peak amplitude was averaged from 25 pulses on a given slice and normalized to baseline peak averages. Dots in (A,C,E) represent individual slices. Sample size is noted on x-axis, with slice number followed by mouse number in parentheses. Larger dots represent mean +/- 1 standard error. *p<0.0001.

extracellular calcium (one-sample t-tests normalized to baseline compared to 1.0, N1: zero calcium, t=-0.01, p=0.99, calcium wash-in: t=0.03, p=0.98; N2: zero calcium, t=-22.4, p<0.0001, calcium wash-in, t=-1.5, p=0.19) and was blocked by the AMPA receptor antagonist, DNQX, in a dose-dependent manner (N1: 1μM, t=1.1, p=0.33, 10μM, t=-0.7, p=0.50; N2: 1μM, t=-29.6, p<0.0001, 10μM, t=-58.1, p<0.0001, 1μM vs. 10μM, t=-10.9, p<0.0001), suggesting that glutamate release from synaptic vesicles within IL terminals in response to light drove the N2 component of the waveform, which reflected the excitatory post-synaptic (NAcSh) population potential of the IL-NAcSh circuit (Figure 9.1B, Figure 9.4). This suggested that the earlier N1 component reflected evoked pre-synaptic firing activity in collective IL axons terminating in NAcSh. We found that application of TTX (1μM, voltage-gated sodium channel antagonist) abolished the N2 peak (t=-111.8, p<0.0001) and surprisingly did not abolish the N1 peak but instead reduced it (Figure 9.1B, Figure 9.4, less than 1, t=-10.6, p<0.0001; greater than 0, t=10.5, p<0.0001). This suggests that while IL axons were not capable of eliciting action potentials in the presence of TTX, direct opsin activity was still detectable in IL population fibers but was insufficient to drive terminal synaptic release of glutamate. The timing of such events indicates this bimodal waveform is mono-synaptic.

To determine the extent to which these measurements are sensitive to the induction of long-term plasticity in the IL-NAcSh circuit, we recorded light-evoked population responses from slice preparations before and after applying a stimulation protocol previously shown to induce changes in plasticity in the nucleus accumbens (Figure 9.1E-F, Figure 9.5) (Hearing et al. 2016; Pascoli et al. 2014; 2011; Thomas et al. 2000). Prolonged 10 Hz stimulation (10 min, 4 ms pulses) caused a sustained reduction in peak amplitude of the N2 component of the light-evoked population response (one-sample t-tests normalized to baseline compared to 1.0, t=-6.7, p<0.0001), while having no lasting effect on the N1 component (Figure 9.5, t=0.9, p=0.36). Consistent with past reports, this suggests that the 10 Hz protocol can induce LTD in the IL-NAcSh circuit, decreasing NAcSh responsivity to IL recruitment (Hearing et al. 2016; Pascoli et al. 2014; 2011; Thomas et al. 2000). We tested this against a different stimulation protocol. We found no changes in either the N1 or N2 components following 100 Hz burst stimulations (1 s train, 4 ms pulses, 4 trains, 10 s inter-train-interval, Figure 9.1F, Figure 9.5, N1: t=1.6, p=0.18; N2: t=0.7, p=0.53). This appeared no

Figure 9.5: Validating optogenetic-driven induction of plasticity in IL-NAcSh



Recordings were sampled at regular fixed intervals every 15 s. Peak N1 and N2 amplitudes were calculated from an average of 5 waveforms plotted every 1.25 min. After 12.5 min of recording baseline (50 waveforms yielding 10 average points), at time zero, slices were either exposed to 10min of no stimulation (A-B), a 10 Hz protocol (C-D) or a 100 Hz protocol (E-F). Regular sampling then resumed immediately after protocols ended for an additional 37.5 min (150 waveforms yielding 30 average points). Only the 10 Hz protocol induced long-term depression in the N2 component without eliciting a lasting effect on the N1 component. No changes were observed in either the N1 or N2 components in either the no stimulation or 100 Hz conditions. Sample size is noted above x-axis, with slice number followed by mouse number in parentheses. Plotted dots represent group mean across slices +/- 1 standard error. Example waveform traces plotted in (B,D,F) taken from an average of 50 baseline pulses and of the final 50 pulses. *p<0.0001, not significant (n.s.).

different from slices that received neither 10 Hz nor 100 Hz stimulation (Figure 9.1F, Figure 9.5, N1: t=0.4, p=0.69; N2: t=0.7, p=0.51).

Measuring strength of synaptic transmission of a specific neural pathway that can be compared across subjects has many challenges. Between-subject differences that can alter extracellular voltage readings include virus transfection rate, opsin expression level, slice preparation cut angle and tissue health, density of IL nerve terminals or NAcSh cell bodies, placement of the recording electrode, etc. One means of circumventing these issues to compare the degree of synaptic strength across subjects is to measure evoked AMPAR/NMDAR current ratios with whole-cell patch clamp recordings (Thomas et al. 2001). Thus, while optogenetically evoked AMPAR/NMDAR ratios can also provide input-specific readings of strength of synaptic transmission, these measures reflect activity at only the single-cell level (Hearing et al. 2016). Here, the clear distinction between pre-synaptic and post-synaptic components of the IL-NAcSh circuit in this light-evoked population waveform provides a possibility to normalize the size of NAcSh response (N2) to the degree of IL recruitment (N1). Although the relative sizes of N1 and N2 can change due to the aforementioned issues (e.g., electrode placement), the relative changes between N1 and N2 in response to alterations in stimulation drive should provide an accurate measure of strength of synaptic transmission in this circuit. We tested the input-output (IO) relationship in NAcSh brain slices from mice in different treatment groups by ramping up and ramping down stimulus drive (here, stimulus duration) and measuring peak N1 and N2 amplitudes (Figure 9.1C-D). The slope of the resulting curve yields a metric of functional synaptic strength of the IL-NAcSh circuit that should be comparable across animals (Figure 9.1D).

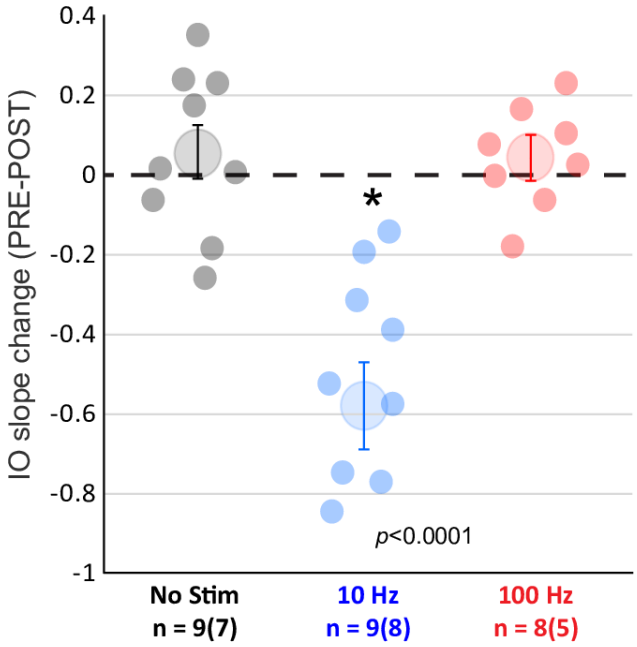
To validate the light-evoked IO relationship as a metric of strength of synaptic transmission in the IL-NAcSh circuit recorded at the population level, we applied the IO assay in slice preparations before and after applying plasticity protocols (no stimulation, 10 Hz, or 100 Hz, Figure 9.1E,G). In each slice, we determined the stimulation value that elicited 50% half-max response in the initial IO assay (mean stimulation duration used for static tests: 0.678 +/- 0.045 ms) and used this value as the test stimulus during the readings before and after the plasticity induction protocol (stimulation used during application of the 10 Hz and 100 Hz protocols

was always 4 ms pulses). The final IO assay always used the same stimulation regimen as the initial IO assay. We found that only the 10 Hz protocol caused a significant decrease in the IO relationship between N1 and N2 response pairs when comparing the final IO assay to the initial IO assay (Figure 9.1G, Figure 9.6, paired t-test, No Stim: t=0.8, p=0.43; 10 Hz: t=-5.3, p<0.0001; 100 Hz: t=1.0, p=0.36). That is, upon additional recruitment of IL activity, less proportional recruitment of NAcSh responsivity was measured following an LTD-inducing stimulation protocol. Furthermore, the degree of change in the IO relationship was significantly correlated with the degree of peak N2 amplitudes change in static readings following plasticity induction relative to baseline static readings (Figure 9.1H, Pearson r=0.876, p<0.001, see Figure 9.1F and Figure 9.5 for depiction of %N2 change during static stimulations). Thus, the IO relationship provides a useful means of measuring strength of synaptic transmission in the IL-NAcSh circuit that is sensitive to a plasticity-inducing stimulation protocol and can be measured at the population level.

To test if this IO assay could capture changes in synaptic strength in response to plasticity-inducing protocol delivered *in vivo*, we delivered no stimulation, 10 Hz, or 100 Hz protocols to mice implanted with optic fibers directed at NAcSh 24 hr prior to sacrificing and *ex vivo* recording (Figure 9.7A). Use of the IO assay revealed that only the 10 Hz protocol group had a significantly reduced IO relationship metric, reflecting successful *in vivo* induction of LTD of the IL-NAcSh circuit (Figure 9.7B, one-way ANOVA, F=8.0, p<0.01, post-hoc Tukey comparisons: No Stim vs. 10 Hz: t=3.3, p<0.01; No Stim vs. 100 Hz: t=-0.5, p=0.88; 10 Hz vs. 100 Hz: t=-3.7, p<0.01). This confirms that an absolute metric of IL-NAcSh strength of synaptic transmission comparable between subjects can be measured at the neuronal population level, and that the 10 Hz manipulation decreases synaptic strength *in vivo*.

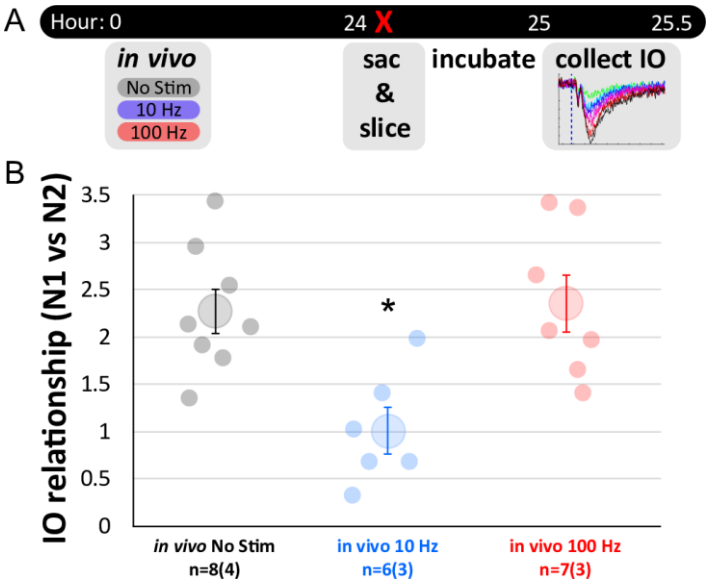
In order to test the role of strength of synaptic transmission in the IL-NAcSh circuit in complex valuation processes, following virus transfection of IL and optic fiber implantation in NAcSh, a separate cohort of mice were trained on a novel variant of a neuroeconomic decision-making task [Restaurant Row(Steiner and Redish 2014; Sweis et al. 2018b)] to work for food in a self-paced manner. In this task, food-restricted

Figure 9.6: Optogenetic IO assays ex vivo capture bath-application of plasticity-inducing protocols



Change in input-output assay from Fig.1G. 10hz group IO slope significantly decreased from zero. Dots represent individual slices. Sample size is noted near the x-axis, with number of slices followed by number of mice in parentheses. Larger dots represent mean +/- 1 standard error.

Figure 9.7: In vivo delivery of optogenetic plasticity protocols measured ex vivo 24hr later



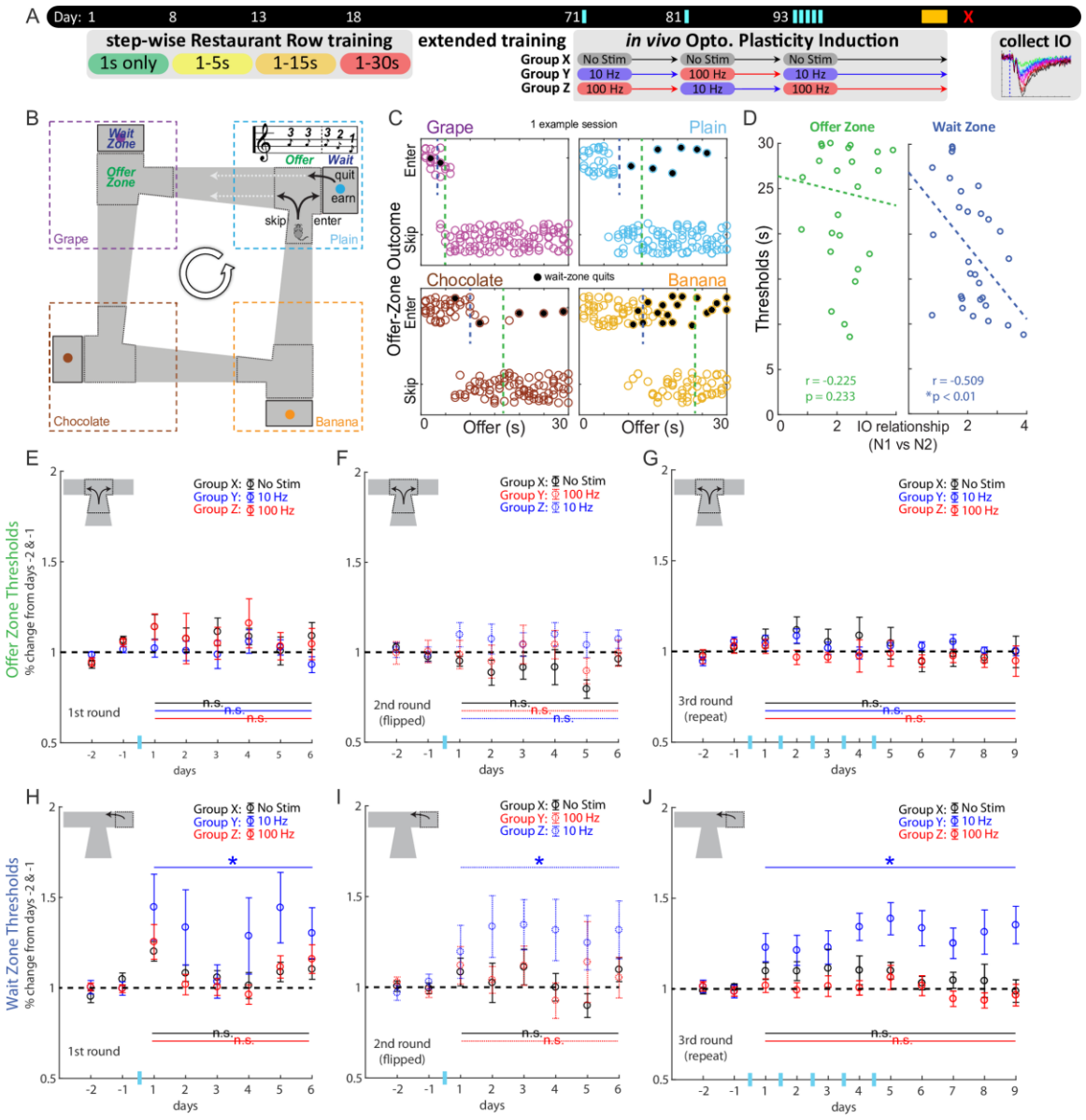
(A) Experimental timeline. Mice were given either no stimulation, 10 Hz, or 100 Hz protocols in vivo at time zero. After 24hr, mice were sacrificed and slices were prepared. Following slice incubation recovery, IO regimen was assayed to compare between group differences in IL-NAcSh strength of synaptic transmission. (B) Only mice receiving the 10 Hz protocol showed a significant decrease in IO slopes measured in the N1 vs. N2 relationship in the IO assay. Dots represent individual slices. Sample size is noted near the x-axis, with number of slices followed by number of mice in parentheses. Larger dots represent mean +/- 1 standard error. *p<0.01.

mice had 1 hr to work for their sole source of food for the day. Thus, this was an economic task in which time must be budgeted to become self-sufficient across days.

Mice traversed a square maze with four feeding sites (restaurants), each with unique spatial cues, each providing a different flavor (Figure 9.8B). On entry into each restaurant, mice were informed of the delay that they would be required to wait to get the food from that restaurant. Mice could then either stay (waiting out the delay) or skip (proceeding on to the next restaurant). Mice revealed preferences for different flavors that varied between animals but were stable across days, indicating subjective valuations for each flavor were used to guide motivated behaviors. Varying flavors, as opposed to varying pellet number, allowed us to manipulate reward value without introducing differences in feeding times (as time is a limited commodity) between restaurants. Flavor preferences were stable across days (Figure 9.9). Costs were measured as different delays mice would have to wait to earn a food reward on that trial, detracting from their session's limited 1 hr time budget. Delays were randomly selected between 1-30 s offers for each trial. Tones sounded upon restaurant entry, with pitch indicating offer cost. Mice learned to deliberate upon offer onset and discriminate between costs based on cued tones, typically rejecting high-cost offers and accepting low-cost offers (Figure 9.10). Taken together, in this task, mice must make serial judgements in a self-paced manner weighing subjective valuations for different flavors against offer costs, balancing the economic utility of sustaining overall food intake against earning more rewards of a desirable flavor. In doing so, cognitive flexibility and self-control become critical components of decision-making valuation processes in this task.

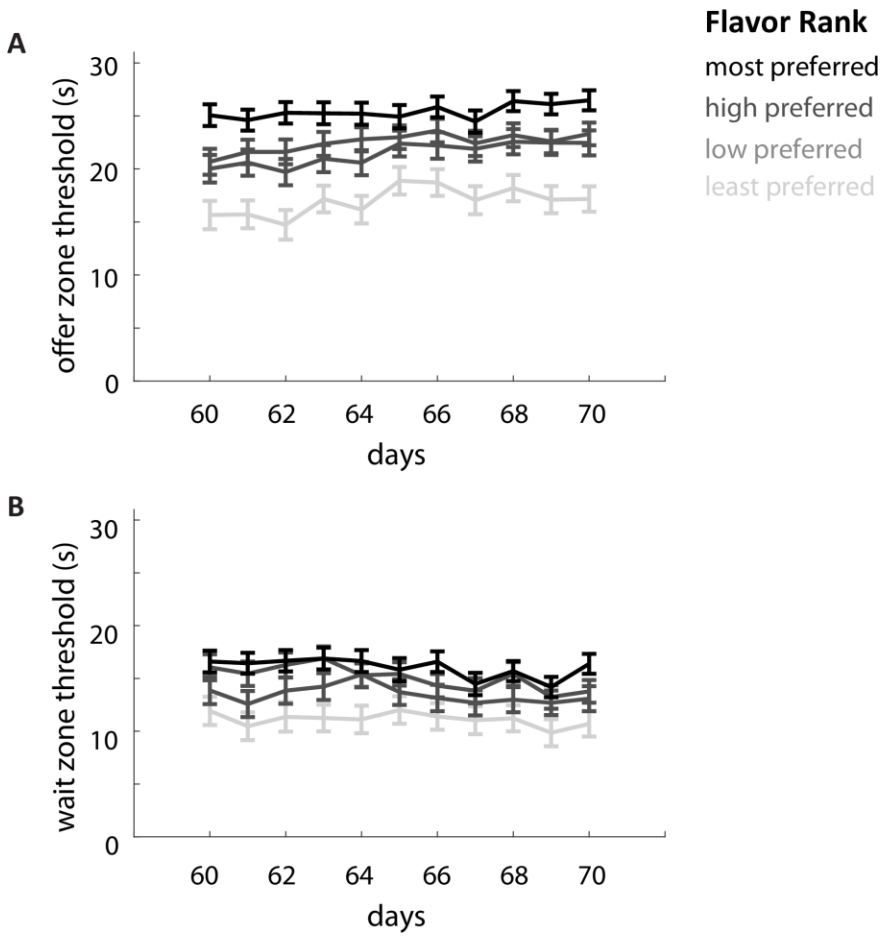
Furthermore, each restaurant contained two distinct zones: an offer zone and a wait zone. Mice were informed of the delay on entry into the offer zone, but delay countdowns did not begin until mice moved into the wait zone. After making an initial enter decision, mice had the opportunity to make a secondary re-evaluative decision to abandon the wait zone (quit) during delay countdowns. Mice were trained in step-wise stages of increasing ranges of offer costs (Figure 9.8A) that initially trained mice to accept the

Figure 9.8: IL-NAcSh plasticity is causally linked to distinct aspects of decision-making valuations



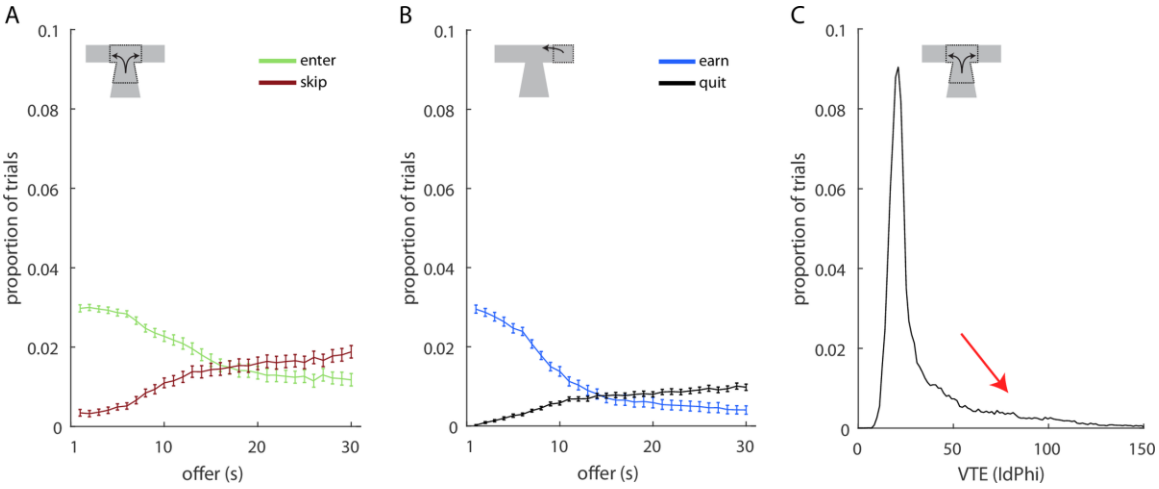
(A) Experimental timeline. 30 mice went through 70 consecutive days of testing in the Restaurant Row task. After day 18, mice were exposed to the full range of 1-30 s offers for the remainder of the experiment. Mice were then divided into 3 groups (X,Y,Z, n=10 each). On the evening of day 71 outside of testing hours, mice received either no stimulation, 10 Hz, or 100 Hz protocols (round 1), and then resumed normal daily Restaurant Row testing. On the evening of day 81, mice received round 2 of stimulation protocols, with groups Y and Z flipping assignments in a repeated measures cross-over design. After day 93, mice received round 3 of stimulation protocols under original group assignments for 5 consecutive evenings. Following behavioral testing, all mice were sacrificed for synaptic strength IO assays. (B) Restaurant Row task schematic. Food-restricted mice were trained on a maze encountering serial offers for flavored rewards in four "restaurants." Restaurant flavor and location were fixed and signaled via contextual cues. Each restaurant contained a separate offer zone and wait zone. Tones sounded in the offer zone; fixed tone pitch indicated delay (1-30 s, randomly selected) mice would have to wait in the wait zone. Tone pitch descended in the wait zone during delay "countdown." Mice could quit the wait zone for the next restaurant during the countdown, terminating the trial. Mice were tested daily for 60 min, earning their only source of food for the day. (C) Example session with individual trials plotted as dots: Mice entered low delays and skipped high delays in the offer zone, while sometimes quitting once in the wait zone (black dots). Dashed vertical lines represent calculated offer zone (green) and wait zone (blue) "thresholds" of willingness to budget time. Thresholds were measured from the inflection point of fitting a sigmoid curve to enters vs. skips or earns vs. quits as a function of delay cost. (D) Only wait zone thresholds significantly correlated (negatively) with synaptic strength of IL-NAcSh while offer zone thresholds did not. Dots represent individual mice. Behavior was taken from final 5 days of testing (gold indicator on timeline in A). (E-J) Only wait zone and not offer zone thresholds significantly changed (increased) following 10 Hz protocol delivery relative to days preceding stimulation in round 1, 2, and 3. Plotted dots represent group mean across mice +/- 1 standard error. *$p<0.0001$ unless otherwise noted, not significant (n.s.).

Figure 9.9: Stability of economic flavor preferences



Offer zone thresholds (A) and wait zone thresholds (B) were calculated for each day in well-trained mice (days 60-70). Flavor preference rankings were determined by calculating total pellets earned in a given restaurant summated across the entire 10 day window and ranked from least to most preferred. Data represent cohort (n=30) mean +/- 1 standard error.

Figure 9.10: Economic choices and deliberative behaviors in the offer and wait zones



Data taken from 10 days of baseline (days 60-70) from all 30 mice. (A) Offer zone choice probability between skipping vs. entering as a function of offer length. Mice entered low delay offers and skipped high delay offers based on cue tone pitch presented at trial onset (restaurant entry). (B) Wait zone choice probability between quitting during the countdown vs. earning as a function of offer length. Plotted lines represent group mean across mice +/- 1 standard error. (C) In the offer zone, mice displayed "pause and look" behaviors, known as "vicarious trial and error" (VTE). VTE reveals on-going deliberation and planning during moments of indecision, supported by numerous electrophysiological experiments. VTE is measured as the absolute integrated angular velocity (in IdPhi units) calculated from the animal's X-Y-location path-trajectory measured from offer onset until either a skip or enter decision is made. The right-tailed shape (red arrow) of the distribution of IdPhi reflect population of trials with increasing deliberative behaviors during moments of indecision.

majority of offers in the offer zone when all costs were relatively inexpensive that they subsequently learned to quit in later stages of training as costs increased.

In order to capture economic valuations of two different stages of decision-making information processing, we calculated separate offer zone "thresholds" of willingness to enter and wait zone "thresholds" of willingness to wait. Sigmoid curves were fit to enter/skip decisions or earn/quit decisions as a function of offer delay and inflection points of each curve were calculated for offer zone and wait zone thresholds, respectively (Figure 9.8C).

On completing their behavioral training, we separated mice into three experimental groups that determined which stimulation protocols they would each receive (Figure 9.8A). As noted above, only the 10 Hz stimulation protocol alters synaptic strength of the IL-NAcSh circuit. All stimulation protocols took place in the evening 4 hr after mice completed testing in Restaurant Row for that day. We also divided the experimental timeline into 3 rounds of stimulation exposure and varied the protocol sequence order and repetition number to allow for between-group as well as within-group comparisons. For the first round of stimulation exposure (Figure 9.8A, 1st cyan timepoint), mice received one evening of their assigned stimulation protocol and were tested as usual in Restaurant Row for the next 10 days. After this 10th day of Restaurant Row testing, for the second round of stimulation exposure (Figure 9.8A, 2nd cyan timepoint), mice received one evening of the opposite stimulation protocol and were tested for an additional 10 days in Restaurant Row. Lastly, for the third round of stimulation exposure (Figure 9.8A, cyan timepoint train), mice received 5 consecutive evenings of their original stimulation protocol assignments and were tested in Restaurant Row for the remaining days of the experiment.

We normalized offer zone and wait zone thresholds to two days of stable testing before each round of stimulation exposure and measured any changes in thresholds relative to baseline over the subsequent days. We found that across all days of testing, regardless of timepoint, no changes were observed in offer zone thresholds in any experimental group (Figure 9.8E-G, repeated measures two-way mixed ANOVA

comparing stim condition x day with mouse as a random variable; round 1: F=0.67, p=0.51; round 2: F=2.46, p=0.11; round 3: F=0.68, p=0.51). However, we found significant increases in wait zone thresholds in the animals receiving the 10 Hz stimulation protocol in each of the three rounds of simulation exposure (Figure 9.8H-J, Table 9.1). This effect was most robust in the third round of stimulation exposure with repeated evenings of stimulation and persisted for several days after the final protocol exposure (Figure 9.8J). Mice displayed no changes in number of laps run, food intake, or locomotion (Figure 9.11, repeated measures two-way mixed ANOVA comparing stim condition x time point with mouse as a random variable, laps: F=1.42, p=0.26; pellets earned: F=0.54, p=0.59; travel time between restaurants: F=1.23, p=0.31). No gross changes in task learning were observed (Figure 9.12). Changes in behavior were specific to valuations made in the wait zone and not the offer zone, with reductions in quit decisions but no change in enter/skip decisions in the 10 Hz group (Figure 9.13, Figure 9.14). Furthermore, LTD of the IL-NAcSh circuit disrupted not only the frequency of quit events but also the economic characteristics of quit decisions, making those re-evaluations less efficient (Figure 9.15). Taken together, this suggests that foraging re-evaluations and self-control capabilities were differentially disrupted independent from initial commitment valuations.

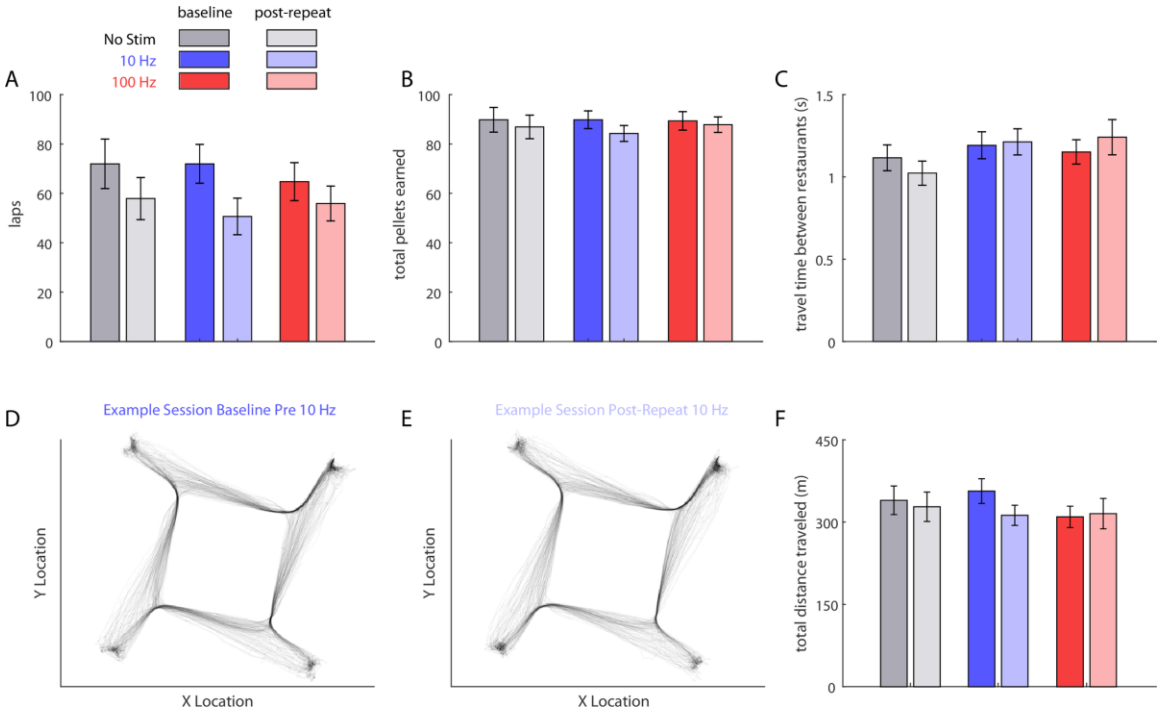
Lastly, all mice were sacrificed and slices were prepared for the *ex vivo* IO assay measuring strength of synaptic transmission of the IL-NAcSh circuit. We calculated IO slopes and correlated them against offer zone and wait zone thresholds over the final 5 days of testing (Figure 9.8A, gold timepoint). We found a significant correlation between synaptic strength of the IL-NAcSh circuit and wait zone thresholds but not offer zone thresholds (Figure 9.8D, correcting for multiple comparisons, Pearson, offer zone: r=-0.225, p=0.23; wait zone: r=-0.509, p<0.01). To ensure that the 10 Hz group (whose behavior was augmented by our plasticity intervention) was not driving this effect, we re-ran correlations excluding this group and found that the correlation between wait zone thresholds and synaptic strength of the IL-NAcSh circuit still held (offer zone: r=-0.325, p=0.16; wait zone: r=-0.499, p<0.05).

Table 9-1: Statistical report of Figure 9.8H-J

Statistical test			Stimulation Round 1	Stimulation Round 2	Stimulation Round 3
ANOVA stim. x days			F=4.58, p<0.05	F=4.10, p<0.05	F=14.05, p<0.0001
	Tukey post-hoc: baseline stim vs. **post no stim**		t=-1.6, p=0.61	t=-0.5, p=0.60	t=-1.7, p=0.56
	Tukey post-hoc: baseline 10 Hz vs. **post 10 Hz**		t=-5.7, p<0.0001	t=-4.22, p<0.0001	t=-7.1, p<0.0001
	Tukey post-hoc: baseline 100 Hz vs. **post 100 Hz**		t=-2.4, p=0.12	t=-1.0, p=0.34	t=0.09, p=0.99
		Tukey post-hoc: **post no stim** vs. **post 10 Hz**	t=-3.3, p<0.05	t=-2.3, p<0.05	t=-3.7, p<0.01
		Tukey post-hoc: **post no stim** vs. **post 100 Hz**	t=-0.8, p=0.96	t=-0.3, p=0.78	t=1.2, p=0.84

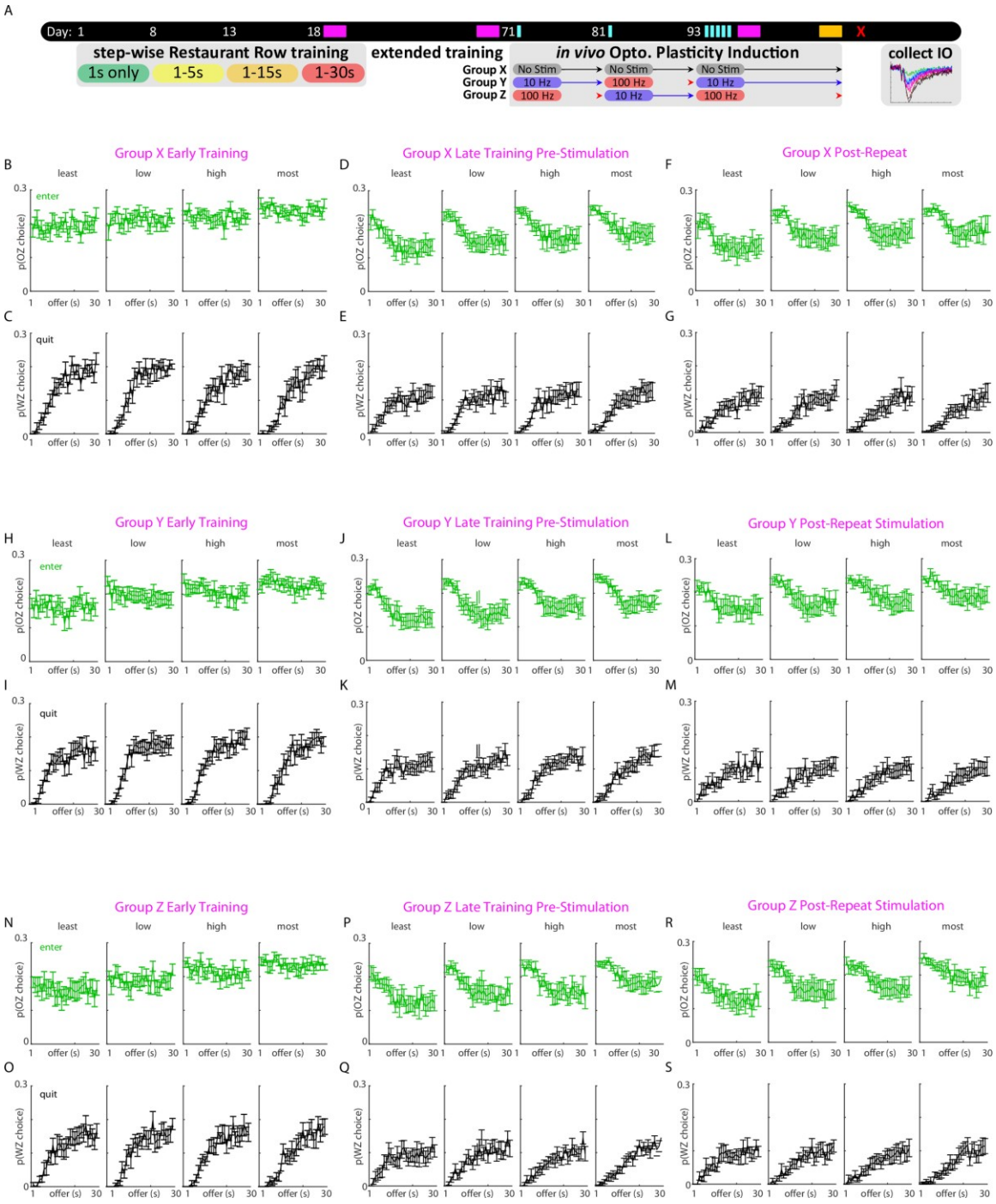
Two-way repeated measures ANOVA was run comparing stimulation condition against time points between stimulation rounds. Interaction effects are reported as well as post-hoc between-group and within-group, across time tests corrected for multiple comparisons. Green text indicates significant findings.

Figure 9.11: Controlling for appetitive or locomotor changes



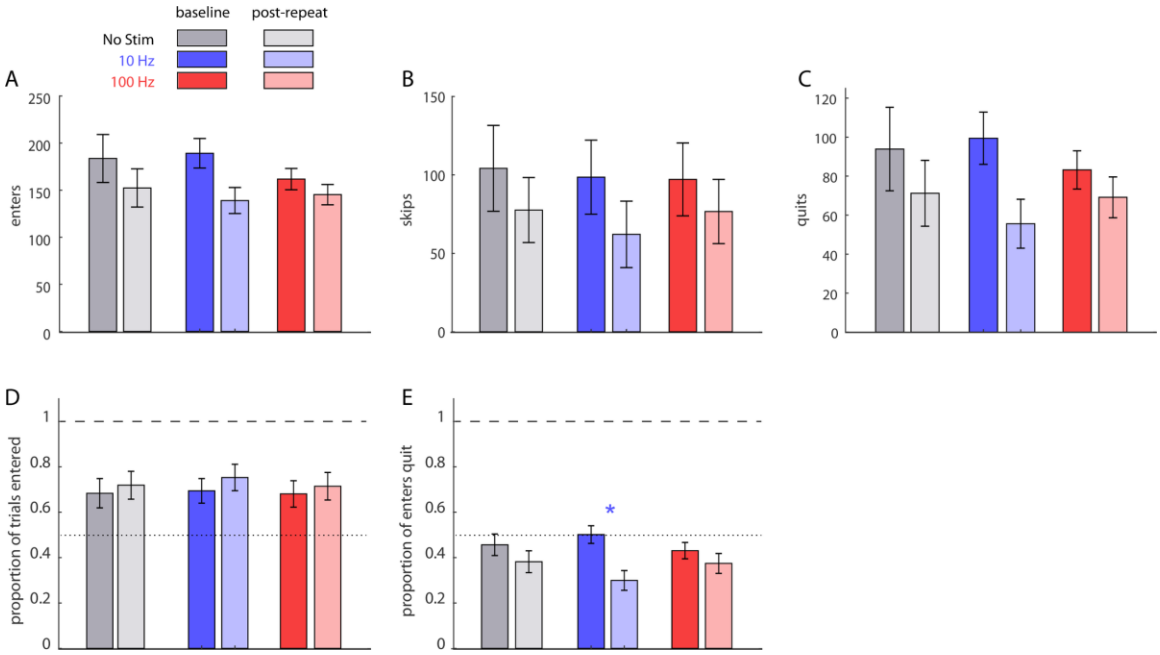
No changes were observed in total number of laps run (A), total number of food pellets earned (B), or locomotor ability (C-F, travel speed or distance traveled) comparing 9 days before the first round of optogenetic stimulation (baseline) and 9 days following round 3 of optogenetic stimulation (post-repeat stimulations, when robust wait zone decision-making changes were observed). Plotted bars represent group mean (n=10 per group) across mice +/- 1 standard error. Track plots in (D-E) depict an example testing session X-Y body tracking positions of a mouse displayed across the entire 1hr session.

Figure 9.12: Controlling for changes in learned spatial task rules



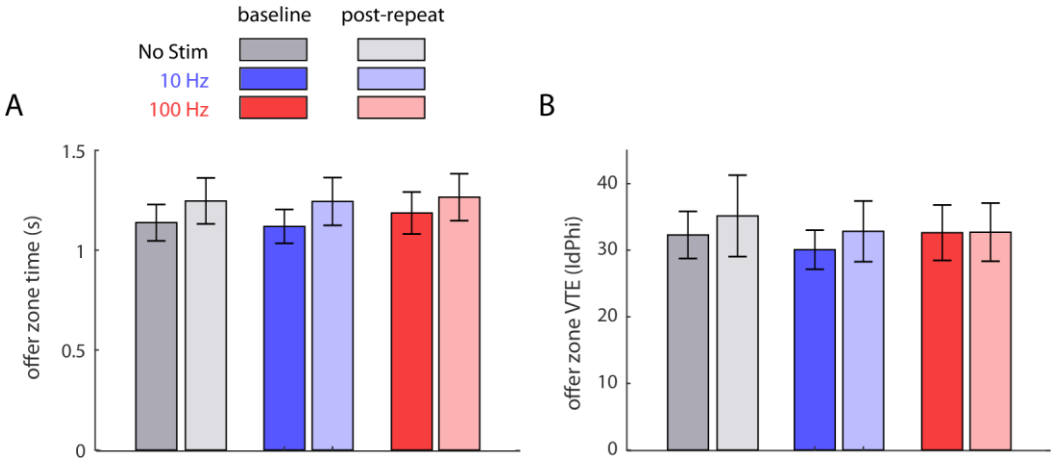
(A) Experimental timeline. Three windows of time marked by magenta blocks indicate early training in the 1-30s environment (days 18-22, B-C,H-I,N-O), late training (days 66-70, D-E,J-K,P-Q), and post-repeat stimulation (days 98-102, F-G,L-M,R-S). (B-S) Offer zone (probability of entering) and wait zone decisions (probability of quitting) depicted as a function of offer cued offer length randomly selected on each trial split by subjective flavor preference rankings (least preferred to most preferred). Experimental treatment conditions are labeled by Groups X, Y, and Z in timeline (A) and throughout (B-S). Early in training (first magenta block, B-C,H-I,N-O), all animals made offer zone decisions to enter depending on subjective flavor preferences (entered more in higher preferred restaurants, $F=11.74$, $p<0.0001$) but with no regard to offer length ($F=1.53$, $p=0.16$). This indicates cued tone information was not incorporated into offer zone decisions and economic behaviors relied solely on wait-zone quit decisions. By late in training (second magenta block, D-E,J-K,P-Q), all animals learned to discriminate randomly presented offer lengths in the offer zone while still ascribing different subjective values to those offers depending on flavor preferences ($F=5.33$, $p<0.01$, restaurant identity communicated through visuospatial cues). Following the stimulation protocols (third magenta block, F-G,L-M,R-S), all animals retained the ability to discriminate randomly presented cued costs in the offer zone while still ascribing subjective value through learned spatial relationships ($F=5.22$, $p<0.01$), but there were no differences between experimental groups ($F=0.29$, $p=0.94$).

Figure 9.13: Characterization of decision types before and after optogenetic manipulations



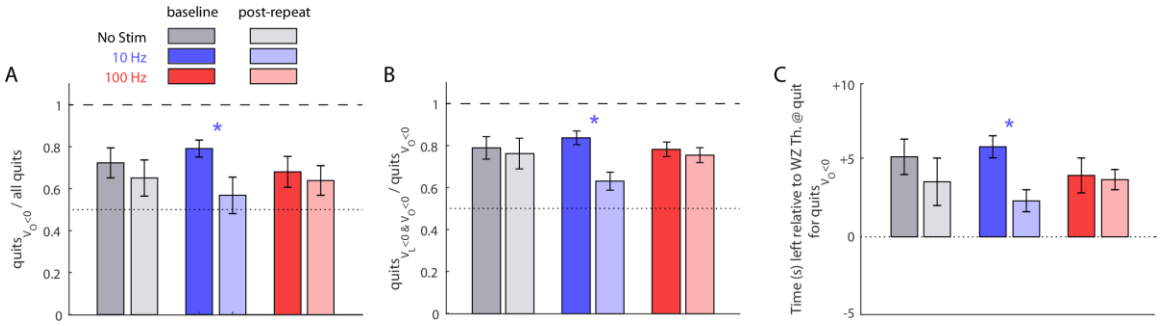
(A-C) No changes were observed in total number enter decisions (A, F=2.11, p=0.14), total number of skip decisions (B, F=0.21, p=0.82), or total number of quit decisions (C, F=2.90, p=0.07). (D) Normalizing number of offer zone decisions to total number of laps run, no changes were observed in proportion of trials that were entered (F=0.15, p=0.86). (E) However, normalizing number of wait zone decisions to total number of offers entered, only the 10 Hz group showed a significant decrease in proportion of enter decisions that were then quit (F=5.94, p<0.01, post-hoc Tukey comparisons: baseline 10 Hz vs. post-repeat 10 Hz, t=6.22, p<0.0001, baseline No Stim vs. post-repeat No Stim, t=2.3, p=0.23, baseline 100 Hz vs. post-repeat 100 Hz, t=1.7, p=0.52). Plotted bars represent group mean (n=10 per group) across mice +/- 1 standard error. *p<0.0001.

Figure 9.14: Characterization of offer zone behaviors before and after optogenetic manipulations



(A-B) No changes were observed in offer zone reaction time (A, F=0.16, p=0.86) or offer zone VTE behavior (B, F=0.28, p=0.76). Plotted bars represent group mean (n=10 per group) across mice +/- 1 standard error.

Figure 9.15: Characterization of wait zone behaviors before and after optogenetic manipulations



We quantified the value (V) of a reward offer (O) normalized to subjective economic preferences for each individual and each restaurant within each session by subtracting the offer length from each respective wait zone threshold (VO=WZ th. – offer). Thus, offers where VO<0 reflect negatively valued offers ("bad" deals) where the delay was greater than what mice were typically willing to wait for and earn. (A) We calculated the proportion of quit events where VO<0 relative to all quits and found that the majority of quits for all mice at baseline were for "bad" deals, correcting offer zone decisions of low value (high delay, high cost) that mice had initially accepted. However, only the 10 Hz group showed a significant reduction ($F=3.84$, $p<0.05$, post-hoc Tukey comparisons: baseline 10 Hz vs. post-repeat 10 Hz, $t=4.57$, $p<0.01$, baseline No Stim vs. post-repeat No Stim, $t=1.5$, $p=0.65$, baseline 100 Hz vs. post-repeat 100 Hz, $t=0.9$, $p=0.94$). (B) We then measured the amount of time left remaining in the countdown at the moment mice made quit decisions and calculated the value of time left (VL) relative to wait zone thresholds (VL=WZ th. – time left). Example: a mouse with a 15 s wait zone threshold accepts a 25 s offer and then quits after 5 s of waiting in the wait zone with 20 s left remaining in that trial's countdown. Thus VO=-10 and VL=-5 on this example trial. Therefore, trials where VO<0 and VL<0 reflect "bad" deals that were initially accepted in the offer zone but were quit in the wait zone while the remaining investment required was still a "bad" deal. We calculated the proportion of "bad" deals that were accepted that were quit in an economically efficient manner where deals were still "bad" when quitting and found that all mice at baseline quit in an economically efficient manner on the majority of "bad" deal quits, truly re-evaluating and changing their minds to correct initial offer zone decisions to enter on these trials. However, only the 10 Hz group showed a significant reduction in economic efficiency of "bad" deal quits ($F=3.39$, $p<0.05$, post-hoc Tukey comparisons: baseline 10 Hz vs. post-repeat 10 Hz, $t=3.68$, $p<0.05$, baseline No Stim vs. post-repeat No Stim, $t=0.5$, $p=0.99$, baseline 100 Hz vs. post-repeat 100 Hz, $t=0.5$, $p=0.98$). (C) Lastly, we calculated the average number of seconds relative to wait zone threshold that were remaining in the countdown at the moment of quitting "bad" offers and found that all mice at baseline, across all trials, quit when the time left was above wait zone thresholds by approximately 5s, consistent with the notion in (B). However, only in the 10 Hz group we found a significant reduction ($F=4.14$, $p<0.05$, post-hoc Tukey comparisons: baseline 10 Hz vs. post-repeat 10 Hz, $t=4.23$, $p<0.01$, baseline No Stim vs. post-repeat No Stim, $t=0.5$, $p=0.99$, baseline 100 Hz vs. post-repeat 100 Hz, $t=1.0$, $p=0.91$). Plotted bars represent group mean (n=10 per group) across mice +/- 1 standard error. *$p<0.01$.

Discussion

Based on recent advancements in neuroeconomics, distinct valuation algorithms are thought to be processed in separable neural circuits (Redish 2013). Thus, it is becoming less clear that reward value is calculated in a common currency through a common neural pathway in the brain (Redish et al. 2008; Rangel et al. 2008). The IL-NAcSh circuit has been suggested to play a modulatory role regulating motivated behaviors distinct from primary reward valuation processes, however separating the two is difficult using traditional behavioral tasks (Barker et al. 2014). Furthermore, circuit interrogation approaches that rely on non-physiological manipulations during on-going behavior obscure interpretations of the functional consequences of synaptic remodeling on endogenous information processing. In this study, I directly tested the role of strength of synaptic transmission in the IL-NAcSh circuit in a novel neuroeconomic task capable of behaviorally separating fundamentally distinct aspects of decision-making information processing.

I found that induction of LTD of glutamatergic IL projections to the NAcSh produced lasting changes in wait-zone valuation processes but not in offer-zone valuation processes, even though they were measured within the same trial. Behaviorally dissociable changes in distinct economic valuation algorithms occurred without affecting other potentially confounding factors, such as disruptions in locomotor capabilities, knowledge of the rules of the task, or ability to remember locations and spatial relationships, none of which changed under the manipulation (Figure 9.11, Figure 9.12) Dissociating multiple valuation processes from these other behavioral processes is difficult to do using standard tasks (e.g., operant place preference, lever pressing, nose poking, Barnes, or Morris mazes) that were not designed for valuation process separations. These standard tasks generally embed instances of distinct valuation algorithms that are overlapping or masked by producing indistinguishable consequences.

The separate offer and wait zones used in the Restaurant Row task allowed us to dissociate principal valuations and re-evaluative processes by separating them into two stages of reward-seeking decision making on every trial. What makes this task economic in nature is that it forces subjects to choose between competing options of varying preferences (flavors) in conflict with varying costs (delays) while hungry and on a limited

time budget. In this paradigm, mice were tasked with making decisions based on cost information randomly provided at the start of each trial (i.e., each entry into a restaurant offer zone). Because mice treat these different tones differently in each restaurant, I know that they have the ability to discriminate cost information and also that different values can be ascribed to the tones based on subject preferences that are updated and acted upon differently on every trial (Figure 9.12) Each trial elicited different sets of discrete actions measurable separately in the offer zone and the wait zone. This separation gave us access to different behavioral computations rooted in fundamentally distinct economic valuations.

In the offer zone, mice displayed behavioral hallmarks of deliberative processes when choosing between competing options before making any investment (Figure 9.10, Redish 2016; Tolman 1939; Muenzinger 1956). In order to invest in the decision, mice had to enter the wait zone. In the wait zone, mice either waited out the cued delay or abandoned an on-going investment. Behavioral and economic models suggest that these two processes in the offer zone and wait zone arise from distinct economic processes of intertemporal choice – deliberative vs. foraging valuation algorithms, respectively – that are thought to be rooted in separable neural circuits (Redish et al. 2008; Hare et al. 2009; Rangel et al. 2008; McClure et al. 2004; Annett 1989; Carter and Redish 2016; Wikenheiser et al. 2013; Sweis et al. 2018b; Stephens et al. 2004; Kolling and Akam 2017; Ainslie 1975; Redish et al. 2016; Charnov 1976; Papale et al. 2012), however causal evidence that these processes arise from separable neural circuits has been lacking. Our data definitively show that strength of synaptic transmission of the IL-NAcSh glutamatergic circuit is causally involved in re-evaluations within the wait zone but not primary valuations in the offer zone.

Studies of extinction and reinstatement behavior in both reward seeking and fear learning support the idea that IL is crucial not for expression of principal (i.e., initial) reward or fear valuations, but rather for acquisition of subsequent extinction learning and maintenance of extinction memory (Barker et al. 2014; Milad and Quirk 2002; Zeeb et al. 2015; Chudasama and Robbins 2003; Sierra-Mercado et al. 2010). Inactivation of IL has no immediate effect on initial reward-seeking behaviors or fear-responses (LaLumiere et al. 2010; Keistler et al. 2015; Lebrón et al. 2004; Quirk et al. 2000; Laurent 2009). However, inactivation

of IL impairs key learning processes such that subjects are unable update regulatory valuations in order to extinguish these behaviors (LaLumiere et al. 2010; Keistler et al. 2015; Lebrón et al. 2004; Quirk et al. 2000; Laurent 2009). Furthermore, inactivation of IL *after* extinction learning takes place can provoke spontaneous reinstatement behaviors by lesioning regulatory processes (Peters et al. 2008). Disconnection experiments and more recent circuit-specific chemogenetic and optogenetic studies have identified the IL projections to the NAcSh or amygdala as more refined pathways implicated in these top-down regulatory processes (for reward and fear, respectively, Do-Monte et al. 2015; Bossert et al. 2012; Augur et al. 2016; Gutman et al. 2017; Kim et al. 2017; LaLumiere et al. 2012). Focusing on reward-related behaviors, inhibition of glutamatergic IL-NAcSh projections time-locked to reinstatement-provoking cues enhanced reinstatement while excitation of these projections during cue presentation reduced reinstatement (Gutman et al. 2017). In IL, extinction learning processes were protein-synthesis-dependent and sensitive to manipulations of neurotrophic factors, cell adhesion molecules, and synaptic plasticity (Pascoli et al. 2014; 2011; Hearing et al. 2016; Gass and Chandler 2013; Barker et al. 2012; Peters et al. 2010; Barker et al. 2015; Ma et al. 2014; Santini et al. 2004). Although with certain differences, the regulatory role of IL is generally in register across both reward-seeking and fear-learning tasks (Peters et al. 2009). Nonetheless, several studies have emphasized key circuit differences in both fear- and reward-related processes. Furthermore, there are reports suggesting neural circuits that are engaged when seeking food in order to emerge from and relieve a food-deprived hunger state may be different from circuits engaged when seeking food while sated either in surplus or luxury (Calhoon et al. 2018; Beyeler et al. 2016; Namburi et al. 2015). The Restaurant Row task captures special economic conflicts in food-deprived states – a critical economic contingency of the task shared across days – between wanting highly desired rewards vs. knowing better to resist costly offers and forage elsewhere.

Studies of behavioral extinction support a useful model of how IL serves a regulatory role modulating valuation processes – a role in which new overriding processes are learned rather than in which old valuation learning is removed or forgotten (Bouton 2004; Rescorla 2001). However, it remains to be determined how principal valuations (e.g., originally learned reward seeking) and regulatory processes (e.g., secondarily

learned overriding processes) might co-exist. How are principal valuations and regulatory valuations integrated together to produce a single behavioral output as measured in the maintenance or reinstatement of extinguished behaviors? How might they be processed independently during on-going decision-making if at odds with each other, regardless if the behavioral output is extinguished or reinstated? It would appear that the weight of these separable processes is critical in determining how they compete with each other in parallel. Our data find that using a novel neuroeconomic approach, principal valuations can be behaviorally segregated from re-evaluative processes characterized as change-of-mind decisions within the same trial. Here, I find that re-evaluative processes, but not principal valuations measured in the same trial, are independently sensitive to off-line changes in strength of synaptic transmission of the glutamatergic IL-NAcSh circuit. Our data implicates a role of IL-NAcSh in top-down control of motivated behavior consistent with previous work, however here, I find that plasticity-augmenting manipulations on a neuroeconomic task reveal this is a neural process separate from and in parallel with on-going principal valuations.

Our findings also provide the first proof-of-principal experiments for a technological advancement in circuit-specific plasticity assessment at the ensemble level. I demonstrate how a novel assay can be used to explain individual differences in distinct aspects of behavior using such circuit-strength metrics for between-subject comparisons.

I was able to measure synaptic strength of the glutamatergic IL-NAcSh circuit at the ensemble-level in optogenetic-evoked recordings prepared *ex vivo*. Measuring strength of synaptic transmission between neurons is a daunting technical challenge. Approaches have historically resorted to electrophysiological recordings of electrical stimulation-evoked post-synaptic responses in order to glean a metric of state of synaptic plasticity. That is, the degree of change of post-synaptic responses elicited using the same electrical stimulation tested before and after a plasticity-inducing intervention has often served as a way to assay plasticity itself. Rather than being limited to measure plasticity merely as a change in response from baseline generally only useful for within-subject comparisons, approaches have made strides in developing single-measurement assays that can capture the strength of synaptic transmission, enabling studies of experience-

dependent forms of plasticity useful for between-subject comparisons. Unfortunately, these assays rely on single-cell-level intracellular recordings generally only measured *ex vivo* using transient bath application of pharmacological agents (e.g., AMPAR / NMDAR ratio metrics, Thomas et al. 2001). Evoked field recordings performed at the population ensemble level are readily accessible *in vivo* yet have been limited to short-windowed within-subject measures of plasticity. Other ensemble measures of plasticity have relied on using oscillatory functional coherence measures, yet this requires multi-site recordings, does not directly reflect strength of synaptic transmission, and is only feasible *in vivo* in an intact whole brain (O'Neill et al. 2013).

Here, I developed a new approach to this problem using optogenetics. Our novel method to measure circuit-specific plasticity in an input-specific manner relies on selective release of glutamate coupled with measurements of relationships between degree of pre-synaptic recruitment and post-synaptic responses while only directly activating afferents without introducing stimulus artifacts. This method provides a novel assay of strength of synaptic transmission in a projection-specific circuit that is comparable across animals without relying on a whole-cell patch clamp approach (Thomas et al. 2001). This demonstrates the first finding of circuit-specific plasticity between two brain regions measured at the ensemble level that relies on only a single measurement and enables between-subject comparisons. This assay revealed that stronger connections from excitatory IL neurons to NAcSh measured at the ensemble level were found in individuals with increased capabilities to overturn initial principal valuations. Strength of the IL-NAcSh circuit varied independently of individual differences in principal valuations.

In this study, pathway specificity lies in the identity of the pre-synaptic neuron. Future studies may be able to take advantage of combinatorial and intersection conditional genetic approaches to restrict opsin expression in afferents onto specific post-synaptic neuron sub-populations. For example, several studies have demonstrated functional differences in sub-populations of NAcSh medium spiny neurons depending on their receptor expression profile (e.g., dopamine D1 vs. D2 receptors, Hearing et al. 2016; Calipari et al. 2016; Keeler et al. 2014). Thus, it is possible that ensemble-level measures as well as augmentations of plasticity of glutamatergic afferents from IL exclusively onto one of these two NAcSh sub-populations may give rise

to further circuit specialization. Red-shifted excitatory opsins, for example, could be used to assay strength of synaptic transmission from a second input (e.g., glutamatergic hippocampal projections) into NAcSh with subsequent red-light assays. Although demonstrated here *ex vivo*, future studies will be able to take advantage of this assay and leverage the many cross-sectional and multi-circuit tools made available through optogenetic approaches both *in vivo* and *ex vivo* using the same methodology. By applying this recording assay with optogenetic-evoked local field potentials *in vivo*, future studies would be able to measure circuit-specific strength of synaptic transmission in animals longitudinally using repeated assays.

The 10 Hz stimulation protocol I used has been previously shown to induce LTD in the IL-NAcSh circuit (Thomas et al. 2000; Pascoli et al. 2014; 2011; Hearing et al. 2016; Thomas et al. 2001). The proposed mechanism of this protocol is dependent on extra-synaptic metabotropic glutamate receptors that signal downstream endocytosis of post-synaptic AMPARs, unlike NMDAR-mediated LTD demonstrated with lower frequency protocols (e.g., 1 Hz, Thomas et al. 2000; Pascoli et al. 2014; 2011; Hearing et al. 2016; Thomas et al. 2001). The 100 Hz protocol has been shown to induce long-term potentiation (LTP) in other circuits, however this elicited no change in synaptic strength in our experiment (Thomas et al. 2000; Pascoli et al. 2014; 2011; Hearing et al. 2016; Thomas et al. 2001). Furthermore, induction of LTP in such studies often relied on patching onto the post-synaptic neuron and holding it at a depolarized potential, making measurements at the field level as well as translation to *in vivo* delivery difficult. Therefore, I used burst 100 Hz as a protocol controlling for light exposure and stimulus timing in addition to the no stimulation control group. Furthermore, our repeated measures cross-over design rules out order effects of stimulation protocols. Induction of heterosynaptic plasticity or retrograde action potentials upon IL terminal stimulation in NAcSh are potential issues to consider. However, despite this, our synaptic strength assay of the glutamatergic IL-NAcSh circuit was capable of capturing plasticity manipulations delivered *in vivo* and importantly could explain behavioral individual differences in re-evaluation processes separate from principal valuations in animals with more vs. less potentiated IL-NAcSh synapses.

These findings have significant implications for how specific aspects of decision making can be processed in a projection-specific circuit, gated by strength of synaptic transmission of that circuit. Maladaptive changes in plasticity is often observed in neuropsychiatric disorders characterized by impairments in decision-making and behavioral regulation, including addiction. In this neuroeconomic framework, the concept of self-control becomes much more nuanced as it can have different implications for the separable neural computations contained within distinct decision-making systems whether it be in a deliberative process or a foraging process. This brings into question the "hypo-frontality" model of addiction that characterizes the inability of individuals with weaker corticostriatal connections to regulate maladaptive motivated behaviors (Kalivas and Volkow 2005; Bickel et al 2007). Here, by combining neuroeconomics and off-line manipulations of circuit-specific plasticity, I demonstrate that the strength of connectivity between two brain structures alters separable aspects of valuation processing during on-going behaviors. Individual differences in "hypo-frontality" here in the IL-NAcSh circuit is causally linked to lasting changes in foraging but not deliberative processes measured within the same trial. Thus, the concept of self-control and impulsivity certainly takes on a circuit-computation-specific definition here.

Of course, many circuits beyond the IL-NACsh pathway are certainly thought to change during the time course of addiction (e.g., drugs on board, during acute withdrawal, immediately following a relapse episode). An obvious next question based on the present findings is: what circuits might be involved in deliberative processes separate from foraging processes? Demonstrations of deliberative algorithms encoded in hippocampal-prelimbic and prelimbic-accumbens core pathways (Johnson et al. 2007, Powell and Redish 2014, Padilla-Coreano et al. 2016, Papale et al. 2016), circuits which have also been shown to change in addiction models (Chen et al. 2013) and are largely non-overlapping with the IL-NAcSh pathway, make interrogating such circuits prime candidates of deliberative-specific computational processes for next steps using this combined neuroeconomics and plasticity approach.

In Chapter 7, I discovered disruptions in separable forms of self-control during the conflict between wanting vs. knowing better - in foraging processes in morphine-abstinent mice distinct from deliberation processes in cocaine-abstinent mice. Thus, here, I build off of Chapter 7 and link mechanisms of circuit-specific memory

and plasticity to separable aspects of decision-making processes – those that go awry in one form of addiction but not another. By taking this combined approach in complex behavioral neuroeconomics and circuit-specific manipulations of plasticity, we can more deeply resolve heterogeneous processes that may be involved in fundamentally distinct etiologies of addiction pathogenesis.

Thus, different individuals suffering from different types of addiction could greatly benefit from neuromodulation therapies that induce long-lasting changes in plasticity. Interestingly, this circuit manipulation, induction of long-term depression in the IL-NAcSh pathway, which has been previously shown to change in both cocaine- and morphine-abstinent mice (Thomas et al. 2000, Thomas et al. 2001, Hearing et al. 2018), has also been shown to provoke reinstatement of conditioned place preference in animals exposed to cocaine (Benneyworth et al. 2018) while blocking reinstatement in animals exposed to morphine (Hearing et al. 2016). That is, the same neural manipulation helped prevent an animal model of relapse in a study using one type of drug of abuse while directly causing release in another type of drug of abuse when tested using simple behavioral paradigms that do not necessarily gain access to separable decision processes. This warrants caution and careful consideration of other circuits that may also be changing in different disease models in conjunction with specific targets that are focused on in select studies.

Currently, there is a growing interest in applying neuromodulation therapies in clinical settings, including the use of transcranial magnetic or deep brain stimulation in neuropsychiatric patients. However, little attention is often paid toward appreciating what plasticity changes might be induced by these treatments over time or what different plasticity states exist in two different patients that may appear to have similar behavioral dysfunctions. Given the complex heterogenous cellular architecture and connectivity of the NAcSh, for example, non-specific neuromodulation therapies such as electrical stimulation can give rise to unpredictable plasticity changes. Plasticity manipulations mentioned above in animal models of addiction using simple behavioral tests have yielded conflicting findings sometimes preventing relapse-like behaviors in some cases while provoking them in others. Thus, one becomes wary that interventions intended to provide therapeutic benefit could potentially worsen disease states. Therefore, future translational studies will require careful

interventions tailored to computation-specific dysfunctions that can be revealed by moving beyond simple tests of value in conjunction with gain-altering circuit manipulations. Only then can we begin to understand the functional consequences of either disease-provoked or intervention-induced synaptic remodeling on complex information processing (Redish and Gordon 2017).

– Interlogue –

Chapter 10

Plasticity in circuit-computation-specific valuation algorithms

In the previous chapter, I discovered in mice that off-line optogenetic manipulations of strength of synaptic transmission specifically between excitatory, glutamatergic, pyramidal neurons that project from the infralimbic (IL) sub-region of the pre-frontal cortex to the shell sub-region of the ventral striatum's nucleus accumbens (NAcSh) was capable of augmenting specific valuations in this novel variant of the Restaurant Row task. Induction of long-term depression (LTD) in these specific IL-to-NAcSh synapses delivered outside of behavioral testing produced lasting changes decisions in the wait zone without affecting offer zone decisions.

In chapters two and four, I demonstrated how the conflict between "wanting" vs. "knowing better" is not only accessible in this novel variant of the Restaurant Row task, but that distinct instances of this conflict in a deliberative modality is separately accessible in the offer zone from instances of this conflict in a foraging modality in the wait zone. Furthermore, it is this foraging modality that captures secondary re-evaluative change-of-mind "opt-out" decisions in the wait zone that appear to capture a distinct aspect of self-control separate from a deliberative modality that captures principal "choose-between" initial commitment decisions in the offer zone.

In chapter six, I discussed how simple behavioral paradigms like conditioned place preference or reward self-administration that are better suited to study mechanisms of memory and learning, and not necessarily how separate bits of information stored as distinct memories are uniquely accessed by separate decision-making systems. The IL-NAcSh circuit in particular has been heavily implicated in mechanisms of learning that involve "top-down" control or executive function processes that are recruited to override existing valuations

when new learning or updating takes place. This is thought to be a form of self-control, operationalized in simple behavioral paradigms as extinction learning, when disadvantageous behaviors are updated and originally learned valuations are suppressed. IL-NAcSh strength increases during extinction learning and manipulations of this connection have been shown to alter reinstatement processes. Thus, it is thought that originally learned valuations are not merely removed or forgotten during extinction learning since they re-emerge following reinstatement.

In the previous chapter, I discovered that processes related to self-control in a foraging modality are mediated through strength of the IL-NAcSh circuit that can co-exist in parallel with separate deliberative valuation processes measured within the same trial, not affected by IL-NAcSh manipulations. This suggests that secondary re-evaluations are not simply re-calculations of reward value made through the same circuits that process initial commitment valuations.

By combining a complex neuroeconomic behavioral paradigm that deconstructs stages of decision making within trial with an off-line manipulation of circuit-specific plasticity, the functional consequences of synaptic remodeling on distinct aspects of decision-making information processing can be realized, as different decision modalities accessed information stored as memories in specific circuits differently.

Furthermore, I developed a novel tool that can assay the strength of a specific circuit, and then used that novel assay to explain individual differences in distinct decision systems. Individuals with stronger IL-NAcSh synapses displayed an increased ability to change their minds in the wait zone after accepting expensive offers in the offer zone. Taken together, considering the discoveries made in chapter 2, wait zone change-of-mind re-evaluations reflect new valuation processes that were learned when mice transitioned from a reward-rich environment to a reward-scarce environment when mice were pressed to exhibit self-control. This reinforces the notion that separate self-control processes learned in the offer zone do indeed reflect fundamentally distinct circuit-computation-specific processes that were unaffected by off-line manipulations of IL-NAcSh synapses.

IL, which is a sub-region of the ventromedial prefrontal cortex, is a highly conserved structure across species, including rodents and primates. This area is thought to be analogous to area 25 in non-human primates. Importantly, projections from IL to the ventral striatum, too, are conserved across species. In a number of neuropsychiatric disorders, particularly those characterized by impairments in self-control, time and time again, these corticostriatal connections are thought to be disrupted. Strength of this connection in functional imaging studies in recovering human addicts can predict susceptibility to relapse. In mouse models of addiction, this circuit has been shown to undergo synaptic remodeling during extinction and reinstatement of drug-seeking behavior in simple behavioral paradigms including drug conditioned place preference or drug self-administration.

The vast majority of animal addiction studies rely on simple behavioral tests of value. The general dogma in the field of addiction neuroscience research is that different forms of addiction, including addiction to different substances of abuse, converge on overlapping pathophysiology and produce similar behavioral dysfunctions. However, based on recent theories in neuroeconomics, it has been hypothesized that because addiction is considered to be a neurobiological disorder of learning and memory giving rise to lasting changes in synaptic plasticity. Because different decision-making systems access stored information differently, distinct failure modes in separate aspects of decision-making information processing could give rise to different forms of addiction. Therefore, addiction is the shared symptom of heterogeneous decision-making diseases. In simple behavioral paradigms, drug-induced changes, which in theory could occur in distinct decision systems, are likely masked by overtly similar changes in maladaptive reward-seeking behavior.

In chapter 7, I demonstrated that prolonged abstinence from chronic cocaine exposure produced dissociable, long-lasting disruptions in decision processes distinct from prolonged abstinence from chronic morphine exposure. The disruptions in foraging behaviors observed in the morphine-abstinent mice matched the changes observed in the previous chapter following LTD-induction of the IL-NAcSh circuit.

Taken together, I took a neuromodulation approach using optogenetics in mice in order to directly interrogate the role of a specific neural circuit in distinct aspects of decision-making information processing by specifically manipulating plasticity off-line. I combined our neuroeconomic task sensitive to distinct decision-making systems with off-line plasticity manipulations to alter the gain of specific circuit in order to understand the functional consequences of synaptic remodeling on decision-making information processing. Through this combined approach – using off-line circuit-specific manipulations in behavioral paradigms that move beyond simple tests of value – I discovered new insights into how multiple, parallel decision-making systems can co-exist and access different information stored as memories differently, even within the same trial.

Chapter 11

On computational psychiatry

The collective work presented in this thesis demonstrates how recent mechanistic, theoretical, methodological, and technological advancements in neuroscience, when integrated in innovative interdisciplinary ways, can give rise to novel insights into how we think about brain processes in health and in disease.

Mental health science and clinical practice is presently on the cusp of a paradigm shift. The recent explosion of discoveries made in neuroscience is accelerating at an unprecedented rate. There is, however, a disconnect between the accelerated rate of discoveries in theoretical and fundamental neuroscience and the development of improved psychiatric treatments, which have not kept pace and have instead remained rather stagnant. Although psychiatric patients have benefited from treatments currently available, clinical endeavors in psychiatry have been plagued by uncertainty at multiple levels of practice: including diagnostics, intervention selection, treatment efficacy monitoring, and patient stability maximization, yielding no clear path to reliably successful cures.

The challenges faced in psychiatry that limit the development of improved treatments emphasize the need for *diagnostic nosology* and *new biomarkers* that speak more closely at the level with which the brain operates.

Chapter based on concepts from:

Redish AD, Gordon JA. 2016. Computational Psychiatry: New Perspectives on Mental Illness. MIT Press.

Before one can treat a disorder, one must know what the disorder is. Outward behavioral symptomology, however subjectively reported and interpreted or objectively measured, is the foundation upon which the Diagnostic and Statistical Manual of Mental Disorders (DSM) is built and treatments are based. Yet, in general, this speaks very little to the underlying neural processes that give rise to behavior. Circuits drive behavior. And different neural processes in distinct neural circuits can drive grossly similar, overtly identical behaviors. At the physiological level, linking specific brain processes or physiological states to specific symptoms has been confounded by many factors, including a lack of understanding of how these circuits are wired together, how circuits perform distinct behavioral computations, and how these circuits can change over time with learning or insult. Biomarkers proposed in the past based on disease categorization rooted in DSM classifications often fail to prove useful in clinical applications when guiding treatment responsivity or disease prognosis and trajectory. Psychiatric diseases are complex, heterogeneous, and often comorbid with each other. Basing disease classification and segregation purely on symptomology is problematic.

A computational perspective affords a fresh lens and rigorous theoretical approach to link underlying neural mechanisms to information processing. Understanding how information is processed through specific circuits in the brain and changed as a function of experiences imposed on those circuits can reveal much about the heterogeneity of neural computations that give rise to distinct aspects of both adaptive and maladaptive cognition and behavior.

The work presented in this thesis is rooted in such approach, basing novel experimental design, complex behavioral modeling, and causal circuit-specific manipulations on rigorous theory that functions in a plausible neurobiological framework translatable from non-human animal models across species.

Behavior ultimately is an output of complex decision-making and action-selection processes. A breakdown in any one of a number of heterogeneous serial or parallel neural computations involved in distinct aspects of decision-making information processing can give rise to behavioral dysfunction and psychiatric illness that can appear similar when measured as a singular, simple behavioral output. Illnesses classified instead by

disruptions in distinct neurally-driven behavioral computations can reveal more about the etiology of disease pathology and guide interventions that are more accurately directed to ameliorate neural dysfunctions tailored to the computational process gone awry within an individual. Thus, there is a strong impetus to resolve the computational complexities of decision-making information processing; for, such framework has significant implications for improving our understanding of and treatments for psychiatric illnesses when distinct computations malfunction.

Simple behavior is not enough to reveal underlying neural computations driving those behaviors. Behavior can be better. In this thesis, I demonstrate how basing experimental design and manipulations on rigorous theories in neuroeconomics can be extremely fruitful in this regard. By moving beyond simple tests of value, information representations that are theoretically thought to be circuit-specific can be revealed through complex behavioral analyses. Such an approach can make predictions about specific behaviors in a neuroeconomic framework when tasks are designed based on theoretically distinct neural computational processes.

The first half of this thesis makes the case for how neuroeconomics can help resolve the complexities of decision-making information processing through careful and rigorous analyses of behavioral computations that may more closely reflect fundamentally distinct neural processes that underlie those behaviors.

The second half of this thesis demonstrates the utility of this approach in a disease-relevant manner and directly tested with circuit-specific manipulations of neural plasticity.

I demonstrated how hidden properties of psychiatric illness can be revealed and resolved in non-human animal models of disease. I used addiction as a case-study, where little is known about how neural changes induced by drugs of abuse give rise to distinct disruptions in computations that underlie aspects of decision-making information processing. Different types of addictions are classically, and even to this day, often lumped together as a common disorder of maladaptive reward-seeking behavior with shared underlying

pathology rooted in a neurobiological disorder of learning and memory. By pushing this concept deeper, I demonstrate the utility in appreciating the intimate link between memory and decision making, the heterogeneity of circuit-specific computations, and why these relationships should be considered when informing the types of manipulations used in experimental paradigms that move beyond simple tests of value.

I demonstrated that behavioral changes following chronic exposure to and abstinence from two different classes of drugs of abuse revealed disruptions in fundamentally distinct aspects of decision-making information processing even though these models have been shown to precipitate similar changes in neural plasticity.

Lastly, I demonstrated how interrogating a specific circuit by directly manipulating plasticity is capable of disrupting distinct behavioral computations in a neuroeconomic framework, suggesting that the driver of disease etiology in one form of addiction but not another can perhaps be traced to a specific circuit change.

Taken together, by moving beyond simple tests of value and basing research efforts on rigorous neuroscience theory in a translatable framework, a richer understanding of brain disease etiology that is linked to circuit-computation-specific processes can be revealed. The work in this thesis provides an interdisciplinary demonstration of a path forward in order to develop better diagnostic nosology and disease biomarkers in efforts to bridge new fundamental discoveries in neuroscience with improvements in clinical practice in psychiatry.

Bibliography

Abe H, Lee D. 2011. Distributed Coding of Actual and Hypothetical Outcomes in the Orbital and Dorsolateral Prefrontal Cortex. Neuron 70(4): 731-741.

Abraham WC, Bear MF. 1996. Metaplasticity: the plasticity of synaptic plasticity. Trends Neurosci 19(4): 126-30.

Abram S, Breton Y-A, Schmidt B, Redish AD, MacDonald A. 2016. The Web-Surf Task: A translational model of human decision-making. Cognitive, Affective & Behavioral Neuroscience 16: 37–50.

Ahmed S, Lenoir M, Guillem K. 2013. Neurobiology of addiction versus drug use driven by lack of choice. Current Opinion in Neurobiology 23: 581–587.

Ahmed S. 2005. Imbalance between drug and non-drug reward availability: A major risk factor for addiction. European Journal of Pharmacology 526: 9–20.

Ahmed S. 2010. Validation crisis in animal models of drug addiction: Beyond non-disordered drug use toward drug addiction. Neuroscience & Biobehavioral Reviews 35: 172–184.

Ahn H, Picard R. 2005. Affective-cognitive learning and decision making: A motivational reward framework for affective agents. Affective Computing and Intelligent Interaction 866–87.

Ainslie G. 1975. Specious reward: a behavioral theory of impulsiveness and impulse control. Psychol Bull 82(4): 463-96

Alcantara, A.A., Lim, H.Y., Floyd, C.E., Garces, J., Mendenhall, J.M., Lyons, C.L., Berlanga, M.L. 2011. Cocaine- and morphine-induced synaptic plasticity in the nucleus accumbens. Synapse 65, 309-320.

Amemiya S, Redish AD. 2016. Manipulating Decisiveness in Decision Making: Effects of Clonidine on Hippocampal Search Strategies. The Journal of Neuroscience 36: 814–27.

Anthony J, Warner L, Kessler R. 1994. Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. Exp Clin Psychopharm 2: 244.

Arantes J, Grace R. 2008. Failure to obtain value enhancement by within-trial contrast in simultaneous and successive discriminations. Learning & Behavior 36: 1–11.

Araos P, Pedraz M, Serrano A, Lucena M, Barrios V, García-Marchena N, Campos-Cloute R, Ruiz J, Romero P, Suárez J, et al. 2015. Plasma profile of pro-inflammatory cytokines and chemokines in cocaine users under outpatient treatment: influence of cocaine symptom severity and psychiatric co-morbidity. Addict Biol 20: 756–772.

Arkes H, Ayton P. 1999. The sunk cost and Concorde effects: Are humans less rational than lower animals? Psychological Bulletin 125: 591.

Arkes H, Blumer C. 1985. The psychology of sunk cost. Organizational Behavior and Human Decision Processes 35: 124–140.

Atwood B, Kupferschmidt D, Lovinger D. 2014. Opioids induce dissociable forms of long-term depression of excitatory inputs to the dorsal striatum. Nature Neuroscience 17: 540–548.

Augur, Wyckoff, Aston-Jones, Kalivas, Peters. 2016. Chemogenetic Activation of an Extinction Neural Circuit Reduces Cue-Induced Reinstatement of Cocaine Seeking. Journal of Neuroscience 36(39):10174–10180.

Aw J, Vasconcelos M, Kacelnik A. 2011. How costs affect preferences: experiments on state dependence, hedonic state and within-trial contrast in starlings. Animal Behaviour 81: 1117–1128.

Badiani A, Belin D, Epstein D, Calu D, Shaham Y (2011) Opiate versus psychostimulant addiction: the differences do matter. Nat Rev Neurosci 12:685–700.

Balleine BW, Delgado MR, Hikosaka O. 2007. The role of the dorsal striatum in reward and decision-making. J Neuroscience 27(31): 8161-8165.

Balleine BW, Dickinson A. 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37(4-5): 407-19.

Barker J, Taylor J, Chandler J. 2014. A unifying model of the role of the infralimbic cortex in extinction and habits. Learning & Memory 21: 441–448.

Barker J, Taylor J, Vries T, Peters J. 2015. Brain-derived neurotrophic factor and addiction: Pathological versus therapeutic effects on drug seeking. Brain Research 1628: 68–81.

Barker J, Torregrossa M, Taylor J. 2012. Low prefrontal PSA-NCAM confers risk for alcoholism-related behavior. Nature Neuroscience 15(10): 1356–8.

Barnes CA, Nadel L, Honig WK. Spatial memory deficit in senescent rats. Canadian Journal of Psychology 34(1): 29.

Becker, J., Kieffer, B., Le Merrer, J. 2017. Differential behavioral and molecular alterations upon protracted abstinence from cocaine versus morphine, nicotine, THC and alcohol. Addict Biol 22(5) 1205-1217.

Belin D, Balado E, Piazza PV, Deroche-Gamonet V. 2009. Pattern of Intake and Drug Craving Predict the Development of Cocaine Addiction-like Behavior in Rats. Biological Psychiatry 65(10): 863-868.

Bell D. 1982. Regret in decision making under uncertainty. Operational Research 30: 961–981.

Benneyworth MA, et al. 2018. Synaptic depotentiation and mGluR5 activity in the nucleus accumbens drive cocaine-primed reinstatement (under review).

Berg R van den, Anandalingam K, Zylberberg A, Kiani R, Shadlen M, Wolpert D. 2016. A common mechanism underlies changes of mind about decisions and confidence. eLife 5: e12192.

Berke JD, Breck JT, Eichenbaum H. 2009. Striatal Versus Hippocampal Representations During Win-Stay Maze Performance. J Neurophys. 101(3): 1575-87.

Berman D, Dudai Y. 2001. Memory extinction, learning anew, and learning the new: dissociations in the molecular machinery of learning in cortex. Science 291(5512):2417-9

Bernheim BD, Rangel A. 2004. Addiction and Cue-Triggered Decision Processes. Am Econ Rev 94: 1558–90.

Berridge K. 1996. Food reward: Brain substrates of wanting and liking. Neuroscience & Biobehavioral Reviews 20: 1–25.

Berridge KC, Robinson TE (2016) Liking, wanting, and the incentive-sensitization theory of addiction. Am Psychol 71:670–679.

Berridge KC, Robinson TE. 1998. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Research Reviews 28(3): 309-369.

Beyeler A, Namburi P, Glober G, Simonnet C, Calhoon G, Conyers G, Luck R, Wildes C, Tye K. 2016. Divergent Routing of Positive and Negative Information from the Amygdala during Memory Retrieval. Neuron 90.

Bickel W, Jarmolowicz D, Mueller, Gatchalian K, McClure S. 2012. Are executive function and impulsivity antipodes? A conceptual reconstruction with special reference to addiction. Psychopharmacology 221: 361–87.

Bickel WK, Miller ML, Yi R, Kowal BP, Lindquist DM, Pitcock JA. 2007. Behavioral and neuroeconomics of drug addiction: Competing neural systems and temporal discounting processes. Drug & Alcohol Dependence, 90: S85-S91.

Blanchard T, Hayden B. 2014. Neurons in Dorsal Anterior Cingulate Cortex Signal Postdecisional Variables in a Foraging Task. The Journal of Neuroscience 34: 646–655.

Bossert J, et al. 2012. Role of Projections from Ventral Medial Prefrontal Cortex to Nucleus Accumbens Shell in Context-Induced Reinstatement of Heroin Seeking. The Journal of Neuroscience 32(14):4982–4991.

Bossert J, Stern A, Theberge F, Marchant N, Wang H-L, Morales M, Shaham Y. 2012. Role of Projections from Ventral Medial Prefrontal Cortex to Nucleus Accumbens Shell in Context-Induced Reinstatement of Heroin Seeking. The Journal of Neuroscience 32: 4982–4991.

Bouton M. 2004 Context and behavioral processes in extinction. Learning & Memory 11(5): 485–94.

Britt J, Benaliouad F, McDevitt R, Stuber G, Wise R, Bonci A. 2012. Synaptic and behavioral profile of multiple glutamatergic inputs to the nucleus accumbens. Neuron 76: 790–803.

Brog JS, Salyapongse A, Deutch AY, Zahm DS. 1993. The patterns of afferent innervation of the core and shell in the "Accumbens" part of the rat ventral striatum: Immunohistochemical detection of retrogradely transported fluoro-gold. J Comp Neurol. 338(2):255-78

Bruin W, Parker A, Fischhoff B. 2007. Individual differences in adult decision-making competence. Journal of Personality and Social Psychology 92: 938.

Bruin W, Parker A, Fischhoff B. 2012. Explaining adult age differences in decision-making competence. Journal of Behavioral Decision Making 25: 352–360.

Bruin W, Strough J, Parker A. 2014. Getting older isn't all that bad: Better decisions and coping when facing "sunk costs". Psychology and Aging 29: 642.

Burnet B, Connolly KJ. 1981. Gene action and the analysis of behaviour. British Medical Bulletin 37: 107-113

Bushong B, King L, Camerer C, Rangel A. 2010. Pavlovian Processes in Consumer Choice: The Physical Presence of a Good Increases Willingness-to-Pay. American Economic Review 100: 1556–1571.

Byrne R. 2002. Mental models and counterfactual thoughts about what might have been. Trends in Cognitive Sciences 6: 426–431.

Calhoon G, Sutton A, Chang C-J, Libster A, Glober G, Leveque C, Murphy G, Namburi P, Leppla C, Siciliano C, et al. 2018. Acute Food Deprivation Rapidly Modifies Valence-Coding Microcircuits in the Amygdala. bioRxiv 285189.

Calipari E, Bagot R, Purushothaman I, Davidson T, Yorgason J, Peña C, Walker D, Pirpinias S, Guise K, Ramakrishnan C, et al. 2016. In vivo imaging identifies temporal signature of D1 and D2 medium spiny neurons in cocaine reward. Proceedings of the National Academy of Sciences of the United States of America 113: 2726–31.

Calipari E, Godino A, Peck E, Salery M, Mervosh N, Landry J, Russo S, Hurd Y, Nestler E, Kiraly D. 2018. Granulocyte-colony stimulating factor controls neural and behavioral plasticity in response to cocaine. Nature Communications 9: 9.

Camchong J, Macdonald A, Mueller B, Nelson B, Specker S, Slaymaker V, Lim K. 2014. Changes in resting functional connectivity during abstinence in stimulant use disorder: a preliminary comparison of relapsers and abstainers. Drug and alcohol dependence 139: 145–51.

Camille N, Goricelli G, Sallet J, Pradat-Diehl P, Duhamel J, Sirigu A. 2004. The Involvement of the Orbitofrontal Cortex in the Experience of Regret. Science 304(5674): 1167-1170.

Carroll ME, Lac ST. 1993. Autoshaping i.v. cocaine self-administration in rats: effects of nondrug alternative reinforcers on acquisition. Psychopharmacology 110(1-2): 5-12.

Carter EC, Redish AD. 2016. Rats value time differently on equivalent foraging and delay-discounting tasks. J Exp Psychol Gen 145: 1093–101.

Casey BJ, Jones R, Hare T. 2008. The Adolescent Brain. Annals of the New York Academy of Sciences 1124: 111–126.

Charnov E. 1976. Optimal foraging, the marginal value theorem. Theoretical population biology 9: 129–136.

Chen BT, Yau HJ, Hatch C, Kusumoto-Yoshida I, Cho SL, Hopf FW, Bonci A. 2013. Rescuing cocaine-induced prefrontal cortex hypoactivity prevents compulsive cocaine seeking. Nature 496(7445): 359-62

Chevaleyre V, Takahashi KA, Castillo PE. 2006. Endocannabinoid-mediated synaptic plasticity in the CNS. Annu. Rev. Neurosci. 29:37-76.

Chudasama, Robbins T. 2003. Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 23: 8771–80.

Church R, Miller K, Meck W, Gibbon J. 1991. Symmetrical and asymmetrical sources of variance in temporal generalization. Animal Learning & Behavior 19: 207–214.

Churchland A, Kiani R, Shadlen M. 2008. Decision-making with multiple alternatives. Nature neuroscience 11: 693–702.

Clark J, Hollon N, Phillips P. 2012. Pavlovian valuation systems in learning and decision making. Current Opinion in Neurobiology 22: 1054–1061.

Cohen J, Goldberg M. 1970. The Dissonance Model in Post-Decision Product Evaluation. Journal of Marketing Research 7: 315.

Cohen NJ, Squire LR. 1980. Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. Science 210(4466): 207-210.

Coleman R, Gross M, Sargent R. 1985. Parental investment decision rules: a test in bluegill sunfish. Behavioral Ecology and Sociobiology 18: 59–66.

Corbit LH, Balleine BW. 2005. Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. Journal of Neuroscience 25(4): 962-970.

Corbit LH, Balleine BW. 2011. The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. Journal of Neuroscience 31(33): 11786-11794.

Coricelli G, Critchley H, Joffily M, O'Doherty J, Sirigu A, Dolan R. 2005. Regret and its avoidance: a neuroimaging study of choice behavior. Nature Neuroscience 8: 1255-1262

Coricelli G, Rustichini A. 2010. Counterfactual thinking and emotions: regret and envy learning. Phil Trans Royal Soc B. 365(1538):241-7

Couey J, Meredith R, Spijker S, Poorthuis RB. 2007. Distributed network actions by nicotine increase the threshold for spike-timing-dependent plasticity in prefrontal cortex. Neuron. 54(1):73-87

Coutureau E, Killcross S. 2003. Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav Brain Res 146(1-2): 167-74.

Creed M, Pascoli V, Lüscher C. 2015. Refining deep brain stimulation to emulate optogenetic treatment of synaptic pathology. Science 347: 659–664.

Czoty PW, McCabe C, Nader MA. 2005. Assessment of the Relative Reinforcing Strength of Cocaine in Socially Housed Monkeys Using a Choice Procedure. Journal of Pharmacology and Experimental Therapeutics 312(1): 96-102.

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature neuroscience 8(12):1704.

Dawkins R, Carlisle TR. 1976. Parental investment, mate desertion and a fallacy. Nature 262: 131–133.

Day M, Langston R. 2006. Post training NMDA receptor blockade offers protection from retrograde interference but does not affect memory consolidation in the watermaze. Neuroscience 137(1): 19-28.

Dayan P, Balleine BW. 2002. Reward, motivation, and reinforcement learning. Neuron 36(2): 285-298.

Dayan P, Berridge K. 2014. Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. Cognitive, affective & behavioral neuroscience 14: 473–92.

Dayan P, Niv Y, Seymour B, Daw ND. 2006. The misbehavior of value and the discipline of the will. Neural Networks 19(8): 1153-1160.

Dayan P, Niv Y. 2008. Reinforcement learning: The Good, The Bad and The Ugly. Current Opinion in Neurobiology 18: 185–196.

Deroche-Gamonet V, Belin D, Piazza PV. 2004. Evidence for addiction-like behavior in the rat. Science. 305(5686): 1014-7

Deroche-Gamonet V, Piaza PV. 2014. Psychobiology of cocaine addiction: Contribution of a multi-symptomatic animal model of loss of control. Neuropharmacology 76(B): 437-449.

Dezfouli, A., Piray, P., Keramati, M. M., Ekhtiari, H., Lucas, C., & Mokri, A. (2009). A neurocomputational model for cocaine addiction. Neural computation 21(10):2869-2893.

Dhami M, Schlottmann A, Waldmann M. 2011. Judgment and Decision Making as a Skill. Cambridge University Press

Dickhaut, Rustichini. 2009. A neuroeconomic theory of the decision process. PNAS 106(52): 22145-22150

Dickinson A, Wood N, Smith JW. 2002. Alcohol seeking by rats: action or habit? The Quarterly Journal of Experimental Psychology 55: 331-348.

Do-Monte F, Manzano-Nieves G, Quiñones-Laracuente K, Ramos-Medina L, Quirk G. 2015. Revisiting the Role of Infralimbic Cortex in Fear Extinction with Optogenetics. The Journal of Neuroscience 35: 3607–3615.

Doya K. 1999. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural Netw 12(7-8): 961-974.

Dudai Y, Jan YN, Byers D, Quinn WG, Benzer S. 1976. Dunce, a mutant of Drosophila deficient in learning. Proceedings of the National Academy of Sciences, 73: 1684-1688.

Dudai Y. 1979. Behavioral plasticity in a Drosophila mutant, dunce. Journal of Comparative Physiology A, 130: 271-275.

Eichenbaum H, Cohen NJ. 2014. Can we reconcile the declarative memory and spatial navigation views on hippocampal function? Neuron 83(4): 764-770.

Eichenbaum H, Stewart C, Morris RG. 1990. Hippocampal representation in place learning. J Neurosci 10(11): 3531-42.

Epstude K, Roese N. 2008. The functional theory of counterfactual thinking. Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc 12: 168–92.

Euston DR., Gruber AJ, McNaughton BL. The role of medial prefrontal cortex in memory and decision making. Neuron 76(6): 1057-1070.

Everitt BJ, Robbins TW. 2005. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nature Neuroscience, 8(11):1481.

Everitt BJ. 2014. Neural and psychological mechanisms underlying compulsive drug seeking habits and drug memories – indications for novel treatments of addiction. Eur J Neurosci 40(1): 2163-2182.

Fox H, D'Sa C, Kimmerling A, Siedlarz K, Tuit K, Stowe R, Sinha R. 2012. Immune system inflammation in cocaine dependent individuals: implications for medications development. Hum Psychopharmacol Clin Exp 27: 156–166.

Friedman A, Homma D, Gibb L, Amemori K, Rubin S, Hood A, Riad M, Graybiel A. 2015. A Corticostriatal Path Targeting Striosomes Controls Decision-Making under Conflict. Cell 161: 1320–33.

Frydman C, Camerer C. 2016. Neural evidence of regret and its implications for investor behavior. The Review of Financial Studies 29: 3108–3139.

Gardnder RS, Uttaro MR, Fleming SE, Suarez DF, Ascoli GA, Dumas TC. 2013. A secondary working memory challenge preserves primary place strategies despite overtraining. Learning and Memory 20: 648-656.

Gardner R, Uttaro M, Fleming S, Suarez D, Ascoli G, Dumas T. 2013. A secondary working memory challenge preserves primary place strategies despite overtraining. Learning & Memory 20: 648–656.

Gass, Chandler. 2013. The Plasticity of Extinction: Contribution of the Prefrontal Cortex in Treating Addiction through Inhibitory Learning. Frontiers in psychiatry 4: 46.

Gerhardt S, Liebman JM. 1981. Differential effects of drug treatments on nose-poke and bar-press self-stimulation. Pharmacology Biochemistry and Behavior. 15(5): 767-771

German P, Fields H. 2007a. How Prior Reward Experience Biases Exploratory Movements: A Probabilistic Model. Journal of Neurophysiology 97: 2083–2093.

German P, Fields H. 2007b. Rat Nucleus Accumbens Neurons Persistently Encode Locations Associated With Morphine Reward. Journal of Neurophysiology 97: 2094–2106.

Giancola P, Mezzich A. 2003. Executive functioning, temperament, and drug use involvement in adolescent females with a substance use disorder. Journal of Child Psychology and Psychiatry 44: 857–866.

Giancola P, Tarter R. 1999. Executive Cognitive Functioning and Risk for Substance Abuse. Psychological Science 10: 203–205.

Gilbert D, Ebert J. 2002. Decisions and revisions: The affective forecasting of changeable outcomes. J Pers Soc Psychol 82: 503–14.

Glimcher, Rustichini. 2004. Neuroeconomics: the consilience of brain and decision. Science 306(5659): 447-452

Goldman M.S., Brown S.A., and Christiansen B.A. (1987). Expectancy theory: Thinking about drinking. In Psychological Theories of Drinking and Alcoholism, eds. H.T. Blaine and K.E. Leonard (New York: Guilford), pp. 181–226.

Goriounova N, Mansvelder H. 2012. Nicotine exposure during adolescence leads to short-and long-term changes in spike timing-dependent plasticity in rat prefrontal cortex. Journal of Neuroscience 32(31): 10484-10493.

Gosnell BA. 2000. Sucrose intake predicts rate of acquisition of cocaine self-administration. Psychopharmacology 149(3): 286-292.

Graves J, Micheyl C, Oxenham A. 2014. Expectations for melodic contours transcend pitch. Journal of Experimental Psychology: Human Perception and Performance 40: 2338.

Graybiel AM, Grafton ST. 2015. The striatum: Where skills and habits meet. CSH Perspectives in Biology 7: a021691.

Graybiel AM. 1998. The basal ganglia and chunking of action repertoires. Neurobiology of learning and memory, 70(1-2), 119-136.

Gremel CM, Costa RM. 2013. Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. Nature Communications 4:2264.

Guo Q, Zhou J, Feng Q, Lin R, Gong H. 2015. Multi-channel fiber photometry for population neuronal activity recording. Biomedical Optics. 6(10)

Gupta A, Meer M, Touretzky D, Redish A. 2012. Segmentation of spatial experience by hippocampal theta sequences. Nature Neuroscience 15: 1032–9.

Gutman AL, Nett KE, Cosme CV, Worth WR, Gupta SC, Wemmie JA, LaLumiere RT. 2017. Extinction of cocaine seeking requires a window of infralimbic pyramidal neuron activity after unreinforced lever presses. J Neurosci. 37(25): 6075-6086

Hare T, Camerer C, Rangel A. 2009. Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. Science 324: 646–648.

Harnett M, Bernier B, Ahn K, Morikawa H. 2009. Burst-timing-dependent plasticity of NMDA receptor-mediated transmission in midbrain dopamine neurons. Neuron. 62(6): 826-38

Hayden BY, Nair AC, McCoy AN, Platt ML. 2008. Posterior Cingulate Cortex Mediates Outcome-Contingent Allocation of Behavior. Neuron 60(1): 19-25.

Hayden BY, Pearson JM, Patt ML. 2009. Fictive reward signals in the anterior cingulate cortex. Science 324: 948-950.

Hearing M, Graziane N, Dong Y, Thomas M. 2018. Opioid and Psychostimulant Plasticity: Targeting Overlap in Nucleus Accumbens Glutamate Signaling. Trends Pharmacol Sci. 39(3): 276-294

Hearing MC, Jedynak J, Ebner SR, Ingebretson A, Asp AJ, Fischer RA, Schmidt C, Larson EB, Thomas M. 2016. Reversal of morphine-induced cell-type–specific synaptic plasticity in the nucleus accumbens shell blocks reinstatement. Proceedings of the National Academy of Sciences 113: 757–762.

Hernandez G, Cheer J. 2015. To Act or Not to Act: Endocannabinoid/Dopamine Interactions in Decision-Making. Front Behav Neurosci 9: 336.

Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, Doya K. 1999. Parallel neural networks for learning sequential problems. Trends Neurosci 22(10): 464-71.

Höffler. 2005. Why humans care about sunk costs while animals don't. An evolutionary explanation. MPI Collective Goods. 17

Holroyd C, Coles M. 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychological review 109: 679–709.

Hull CL. 1952. A Behavior System: An Introduction to Behavior Theory Concerning the Individual Organism. Yale University Press, New Haven.

Hyman S, Malenka R. 2001. Addiction and the brain: The neurobiology of compulsion and its persistence. Nature Reviews Neuroscience 2: 35094560.

Hyman S. 2005. Addiction: A Disease of Learning and Memory. American Journal of Psychiatry 162: 1414–1422.

Jaffe A, Pham J, Tarash I, Getty SS, Fanselow MS, Jentsch DJ. 2014. The Absence of Blocking Innicotine High-Responders as a Possible Factor in the Development of Nicotine Dependence? The Open Addiction Journal 7.

John WS, Banala AK, Newman AH, Nader MA. 2015. Effects of buspirone and the dopamine D3 receptor compound PG619 on cocaine and methamphetamine self-administration in rhesus monkeys using a food-drug choice paradigm. Psychopharmacology 232(7): 1279-1289.

Johnson A, Redish A. 2007. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. J Neurosci 27: 12176–89.

Johnson A, van der Meer MA, Redish AD. 2007. Integrating hippocampus and striatum in decision-making. Current opinion in neurobiology 17: 692–7.

Kacelnik A, Marsh B. 2002. Cost can increase preference in starlings. Animal Behaviour 63: 245–250.

Kahneman D, Knetsch J, Thaler R. 1991. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. Journal of Economic Perspectives 5: 193–206.

Kalenscher T, van Wingerden M. 2011. Why we should use animals to study economic decision making - a perspective. Front Neurosci 5: 82.

Kalivas PW, Volkow ND. 2005. The neural basis of addiction: a pathology of motivation and choice. American Journal of Psychiatry, 162(8): 1403-1413.

Karlsson C, Schank J, Rehman F, Stojakovic A, Björk K, Barbier E, Solomon M, Tapocik J, Engblom D, Thorsell A, et al. 2017. Proinflammatory signaling regulates voluntary alcohol intake and stress-induced consumption after exposure to social defeat stress in mice. Addiction Biology 22: 1279–1288.

Kasanetz F, Deroche-Gamonet V, Berson N, Balado E, Lafourcade M, Manzoni O, Piazza PV. 2010. Transition to addiction is associated with a persistent impairment in synaptic plasticity. Science 328: 1709–12.

Kauer JA, Malenka RC. 2007. Synaptic plasticity and addiction. Nat Rev Neurosci 8: 844–58.

Keeler JF, Pretsell DO, Robbins TW. 2014. Functional implications of dopamine D1 vs. D2 receptors: A 'prepare and select' model of the striatal direct vs. indirect pathways. Neuroscience 282: 156–175.

Keistler C, Barker J, Taylor J. 2015. Infralimbic prefrontal cortex interacts with nucleus accumbens shell to unmask expression of outcome-selective Pavlovian-to-instrumental transfer. Learning & memory 22: 509–13.

Kermer D, Driver-Linn E, Wilson T, Gilbert D. 2006. Loss aversion is an affective forecasting error. Psychol Sci 17: 649–53.

Kim C, Adhikari A, Deisseroth K. 2017a. Integration of optogenetics with complementary methodologies in systems neuroscience. Nature Reviews Neuroscience. 18(4): 222-235

Kim C, Ye L, Jennings J, Pichamoorthy N, Tang D, Yoo A-C, Ramakrishnan C, Deisseroth K. 2017b. Molecular and Circuit-Dynamical Identification of Top-Down Neural Mechanisms for Restraint of Reward Seeking. Cell 170: 1013-1027.e14.

Kim H, Shimojo S, O'Doherty J. 2006. Is Avoiding an Aversive Outcome Rewarding? Neural Substrates of Avoidance Learning in the Human Brain. PLoS Biology 4: e233.

Kim J, Jung M. 2006. Neural circuits and mechanisms involved in Pavlovian fear conditioning: a critical review. Neurosci Biobehav Rev 30: 188–202.

Klaczynski P. 2001a. Analytic and Heuristic Processing Influences on Adolescent Reasoning and Decision-Making. Child Development 72: 844–861.

Klaczynski P. 2001b. Framing effects on adolescent task representations, analytic and heuristic processing, and decision making Implications for the normative/descriptive gap. Journal of Applied Developmental Psychology 22: 289–309.

Klaczynski P. 2004. A dual-process model of adolescent development: implications for decision making, reasoning, and identity. Advances in child development and behavior 32: 73–123.

Klaczynski P. 2009. In two minds: Dual processes and beyond.

Klapoetke N, Murata Y, Kim S, Pulver S, Birdsey-Benson A, Cho Y, Morimoto T, Chuong A, Carpenter E, Tian Z, et al. 2014. Independent optical excitation of distinct neural populations. Nature methods 11: 338–46.

Knutson B, Greer S. 2008. Anticipatory affect: neural correlates and consequences for choice. Philosophical Transactions of the Royal Society B: Biological Sciences 363: 3771–3786.

Ko D, Wanat M. 2016. Phasic Dopamine Transmission Reflects Initiation Vigor and Exerted Effort in an Action- and Region-Specific Manner. The Journal of neuroscience : the official journal of the Society for Neuroscience 36: 2202–11.

Kodangattil J, Dacher M. 2013. Spike timing-dependent plasticity at GABAergic synapses in the ventral tegmental area. The Journal of Physiology 591(19): 4699-710.

Kolling N, Akam T. 2017. (Reinforcement?) Learning to forage optimally. Current opinion in neurobiology 46: 162–169.

Krypotos A, Effting M, Kindt M, Beckers T. 2015. Avoidance learning: a review of theoretical models and recent developments. Front Behav Neurosci 9: 189.

Kurth-Nelson Z, Bickel WK, Redish AD. 2012. A theoretical account of cognitive effects in delay discounting. European Journal of Neuroscience 35:1052-1064.

Kurth-Nelson Z, Redish AD. 2012. "Modeling decision-making systems in addiction" in Computational Neuroscience of Drug Addiction. B. Gutkin, S. Ahmed (eds). Springer. Chapter 6, pages 163-188.

Lacagnina MJ, Rivera PD, Bilbo SD. 2017. Glial and Neuroimmune Mechanisms as Critical Modulators of Drug Use and Abuse. Neuropsychopharmacology 42: 156–177.

Lakshminaryanan V, Chen M, Santos L. 2018. Endowment effect in capuchin monkeys. Philosophical transactions of the Royal Society of London Series B, Biological sciences 363: 3837–44.

LaLumiere R, Niehoff K, Kalivas P. 2010. The infralimbic cortex regulates the consolidation of extinction after cocaine self-administration. Learning & Memory 17(4):168–175.

LaLumiere R, Smith K, Kalivas P. 2012. Neural circuit competition in cocaine-seeking: roles of the infralimbic cortex and nucleus accumbens shell. European Journal of Neuroscience 35: 614–622.

Laurent V, Westbrook RF. 2009. Inactivation of the infralimbic but not the prelimbic cortex impairs consolidation and retrieval of fear extinction. Learn Mem. 16(9): 520-9.

Lebrón K, Milad M, Quirk G. 2004. Delayed recall of fear extinction in rats with lesions of ventral medial prefrontal cortex. Learning & memory (Cold Spring Harbor, NY) 11: 544–8.

LeDoux J, Daw ND. 2018. Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. Nature Reviews Neuroscience.

LeDoux J. 1998. ed. The emotional brain: The mysterious underpinnings of emotional life. Simon & Schuster. New York.

Lee BR, Dong Y. 2011. Cocaine-induced metaplasticity in the nucleus accumbens: silent synapse and beyond. Neuropharmacology 61: 1060–1069.

Levy DJ, Glimcher PW. 2011. Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. J Neuroscience 31(41): 14693-14707.

Lewitus GM, Konefal SC, Greenhalgh AD, Pribiag H, Augereau K, Stellwagen D. 2016. Microglial TNF-α Suppresses Cocaine-Induced Plasticity and Behavioral Sensitization. Neuron 90: 483–491.

Loewenstein G, Rick S, Cohen J. 2008. Neuroeconomics. Annual review of psychology 59: 647–72.

Loomes G, Surgden R. 1982. Regret Theory: An alternative theory of rational choice under uncertainty. Economic Journal 92: 805–824.

Lu L, Shepard J, Hall F, Shaham Y. 2003. Effect of environmental stressors on opiate and psychostimulant reinforcement, reinstatement and discrimination in rats: a review. Neuroscience & Biobehavioral Rev 27(5):457-91.

Lüscher C, Malenka RC. 2011. Drug-evoked synaptic plasticity in addiction: from molecular changes to circuit remodeling. Neuron 69: 650–63.

Ma Y, Lee BR, Wang X, Guo C, Liu L, Cui R, Lan Y, Balcita-Pedicino JJ, Wolf ME, Sesack SR, Shaham Y, Schluter OM, Huang YH, Dong Y. Bidirectional modulation of incubation of cocaine craving by silent synapse-based remodling of prefrontol cortex to accumbens projections. Neuron 83(6): 1453-1467.

Maestripieri D, Alleva E. 1991. Litter defence and parental investment allocation in house mice. Behavioural Processes 23: 223–230.

Magalhães P, White G. 2016. The sunk cost effect across species: A review of persistence in a course of action due to prior investment. Journal of the experimental analysis of behavior 105: 339–61.

Magalhães P, White, Stewart T, Beeby E, Vliet W van der. 2012. Suboptimal choice in nonhuman animals: rats commit the sunk cost error. Learning & behavior 40: 195–206.

Mameli M, Bellone C, Brown M, Lüscher C. 2011. Cocaine inverts rules for synaptic plasticity of glutamate transmission in the ventral tegmental area. Nature Neuroscience. 14(4): 414-6

Mameli M, Lüscher C. 2011. Synaptic plasticity and addiction: learning mechanisms gone awry. Neuropharmacology 61: 1052–9.

Marchiori D, Warglien M. 2008. Predicting Human Interactive Learning by Regret-Driven Neural Networks. Science 319: 1111–1113.

Marks KR, Kearns DN, Christensen CJ, Silberberg A, Weiss SJ. 2010. Learning that a cocaine reward is smaller than expected: A test of Redish's computational model of addiction. Behavioral Brain Research 212(2): 204-207.

Martin-Garcia E, Courtin J, Renault P, Fiancette J, Wurtz H, Simonnet A, Levet F, Herry C, Deroche-Gamonet V. 2014. Frequency of Cocaine Self-Administration Influences Drug Seeking in the Rat: Optogenetic Evidence for a Role of the Prelimbic Cortex. Neuropsychopharmacology 39: 2317-2330.

McClure S, Laibson D, Loewenstein G, Cohen J. 2004. Separate neural systems value immediate and delayed monetary rewards. Science 306: 503–7.

McCoy AN, Platt ML. 2005. Risk-sensitive neurons in macaque posterior cingulate cortex. Nat Neurosci 8: 1220-1227.

McDonald RJ, White, NM, 1993. A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. Behavioral Neuroscience 107(1): 3-22.

Milad M, Quirk G. 2002. Neurons in medial prefrontal cortex signal memory for fear extinction. Nature 420: 70–4.

Milner B, Squire L, Kandel E. 1998. Cognitive Neuroscience and the Study of Memory. Neuron 20: 445–468.

Moal M Le, Koob GF. 2007. Drug addiction: pathways to the disease and pathophysiological perspectives. Eur Neuropsychopharmacol 17: 377–93.

Morales M, Margolis E. 2017. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. Nature Reviews Neuroscience 18: 73–85.

Morris RGM, Garrud P, O'Keefe J. 1982. Place navigation impaired in rats with hippocampal lesions. Nature 297: 681-683.

Morsanyi K, Handley S. 2008. How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. Journal of Experimental Child Psychology 99: 18–36.

Muenzinger K. 1956. On the origin and early use of the term vicarious trial and error (VTE). Psychological bulletin 53: 493–4.

Müller CP, Homberg JR. 2015. The role of serotonin in drug use and addiction. Behavioral Brain Research 277: 146-192.

Nadar MA, Hedeker D, Woolverton WL. 1993. Behavioral economics and drug choice: effects of unit price on cocaine self-administration by monkeys. Drug and Alcohol Dependence 33(2): 193-199.

Nadar MA, Woolverton WL. 1991. Effects of increasing the magnitude of an alternative reinforcer on drug choice in a discrete-trials choice procedure. Psychopharmacology 105(2): 169-74.

Nadar MA, Woolverton WL. 1992. Effects of increasing response requirement on choice between cocaine and food in rhesus monkeys. Psychopharmacology 108(3): 295-300.

Namburi P, Beyeler A, Yorozu S, Calhoon G, Halbert S, Wichmann R, Holden S, Mertens K, Anahtar M, Felix-Ortiz A, et al. 2015. A circuit mechanism for differentiating positive and negative associations. Nature 520: 675–678.

Naqvi NH, Bechara A. 2010. The insula and drug addiction: an interoceptive view of pleasure, urges, and decision-making. Brain Structure and Function, 214(5-6): 435-450.

Navarrete M, Araque A. 2008. Endocannabinoids Mediate Neuron-Astrocyte Communication. Neuron 57: 883–893.

Navarrete M, Araque A. 2010. Endocannabinoids Potentiate Synaptic Transmission through Stimulation of Astrocytes. Neuron 68: 113–126.

Nestler EJ. 2001. Molecular basis of long-term plasticity underlying addiction. Nature Reviews Neuroscience 2: 119-128.

Neuhofer D, Kalivas P. 2018. Metaplasticity at the addicted tetrapartite synapse: A common denominator of drug induced adaptations and potential treatment target for addiction. Neurobiology of Learning and Memory. Epub

Northcutt A, Hutchinson M, Wang X, Baratta M, Hiranita T, Cochran T, Pomrenze M, Galer E, Kopajtic T, Li C, et al. 2015. DAT isn't all that: cocaine reward and reinforcement require Toll-like receptor 4 signaling. Molecular Psychiatry 20: 1525–1537.

O'Neill P-K, Gordon J, Sigurdsson T. 2013. Theta Oscillations in the Medial Prefrontal Cortex Are Modulated by Spatial Working Memory and Synchronize with the Hippocampus through Its Ventral Subregion. The Journal of Neuroscience 33: 14211–14224.

O'Keefe J, Nadel L. 1978. lThe Hippocampus as a Cognitive Map. Oxford University Press, Oxford-New York.

Olmstead MC, Lafond MV, Everitt BJ, Dickinson A. 2001. Cocaine seeking by rats is a goal-directed action. Behav Neurosci 115: 394 –402.

Ostroumov A, Dani JA. 2017. Convergent Neuronal Plasticity and Metaplasticity Mechanisms of Stress, Nicotine, and Alcohol. Annual Review of Pharmacology. 58:547-566

Packard MG, Knowlton BJ. 2002. Learning and memory functions of the basal ganglia. Annual review of neuroscience 25(1): 563-593.

Packard MG, McGaugh JL. 1992. Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: further evidence for multiple memory systems. Behav. Neurosci 106(3): 439-446.

Padilla-Coreano N, Bolkan SS, Pierce GM, Blackman DR, Hardin WD, Garcia-Garcia AL, Spellman TJ, Gordon JA. 2016. Direct Ventral Hippocampal-Prefrontal Input Is Required for Anxiety-Related Neural Activity and Behavior. Neuron 89: 857–866.

Papale A, Stott J, Powell N, Regier P, Redish. 2012. Interactions between deliberation and delay-discounting in rats. Cognitive, affective & behavioral neuroscience 12: 513–26.

Papale AE, Zielinski MC, Frank LM, Jadhav SP, Redish AD. 2016. Interplay between Hippocampal Sharp-Wave-Ripple Events and Vicarious Trial and Error Behaviors in Decision Making. Neuron 92: 975–982.

Parker A, Fischhoff B. 2005. Decision-making competence: External validation through an individual-differences approach. Journal of Behavioral Decision Making 18: 1–27.

Pascoli V, Terrier J, Espallergues J, Valjent E, O'Connor E, Lüscher C. 2014. Contrasting forms of cocaine-evoked plasticity control components of relapse. Nature 509(7501): 459-64.

Pascoli V, Turiault M, Lüscher C. 2011. Reversal of cocaine-evoked synaptic potentiation resets drug-induced adaptive behaviour. Nature 481: 71–75.

Patrick V, Lancellotti M, Demello G. 2009. Coping with non-purchase: Managing the stress of inaction regret. Journal of Consumer Psychology 19: 463–472.

Pattison K, Zentall T, Watanabe S. 2012. Sunk cost: pigeons (Columba livia), too, show bias to complete a task rather than shift to another. Journal of comparative psychology (Washington, DC : 1983) 126: 1–9.

Perry A, Westenbroek C, Becker JB. 2013. The development of a preference for cocaine over food identifies individual rats with addiction-like behaviors. PloS one. 8(11)

Perry JL, Larson EB, German JP, Madden GJ, Carroll ME. 2005. Impulsivity (delay discounting) as a predictor of acquisition of IV cocaine self-administration in female rats. Psychopharmacology 178(2-3): 193-201.

Peters J, Kalivas P, Quirk G. 2009. Extinction circuits for fear and addiction overlap in prefrontal cortex. Learning & memory (Cold Spring Harbor, NY) 16: 279–88.

Peters J, LaLumiere R, Kalivas P. 2008. Infralimbic Prefrontal Cortex Is Responsible for Inhibiting Cocaine Seeking in Extinguished Rats. The Journal of Neuroscience 28(23): 6046–6053.

Peters J, Perea L eppa-, Melendez L, Quirk G. 2010. Induction of fear extinction with hippocampal-infralimbic BDNF. Science (New York, NY) 328: 1288–90.

Peters, J., Kalivas, P. W., & Quirk, G. J. 2009. Extinction circuits for fear and addiction overlap in prefrontal cortex. Learning & Memory, 16(5): 279-288.

Piazza PVV, Deminière J-MM, Maccari S, Mormède P, Le Moal M, Simon H (1990) Individual reactivity to novelty predicts probability of amphetamine self-administration. Behav Pharmacol 1:339–345.

Pickard H, Ahmed S, Foddy B. 2015. Alternative models of addiction. Frontiers in psychiatry. 6:20

Piray P, Keramati MM, Dezfouli A, Lucas C, Mokri A. 2010. Individual differences in nucleus accumbens dopamine receptors predict development of addiction-like behavior: a computational approach. Neural computation 22(9): 2334-2368.

Plassmann H, O'Doherty J, Shiv B, Rangel A. 2008. Marketing actions can modulate neural representations of experienced pleasantness. Proceedings of the National Academy of Sciences 105: 1050–1054.

Pompilio L, Kacelnik A, Behmer S. 2006. State-Dependent Learned Valuation Drives Choice in an Invertebrate. Science 311: 1613–1615.

Powell NJ, Redish AD. 2014. Complex neural codes in rat prelimbic cortex are stable across days on a spatial decision task. Frontiers in Behavioral Neuroscience 8: 00120.

Quinn WG, Gould JL. 1979. Nerves and genes. Nature 278: 19-23

Quinn WG, Sziber PP, Booker R. 1979. The Drosophila memory mutant amnesiac. Nature 277: 212-214

Quirk, Russo, Barron, Lebron. 2000. The role of ventromedial prefrontal cortex in the recovery of extinguished fear. The Journal of neuroscience : the official journal of the Society for Neuroscience 20: 6225–31.

Rangel A, Camerer C, Montague R. 2008. A framework for studying the neurobiology of value-based decision making. Nature Reviews Neuroscience 9: 545–556.

Redish AD, Gordon JA. 2017. Computational Psychiatry: New Perspectives on Mental Illness. MIT Press

Redish AD, Jensen S, Johnson A. 2008. A unified framework for addiction: vulnerabilities in the decision process. The Behavioral and brain sciences 31: 415–37; discussion 437-87.

Redish AD, Johnson A. 2007. A computational model of craving and obsession. Annals of the New York Academy of Sciences 1104: 324-339.

Redish AD, Mizumori SJ. 2015. Memory and decision making. Neurobiology of learning and memory 117: 1–3.

Redish AD, Schultheiss N, Carter E. 2016. The Computational Complexity of Valuation and Motivational Forces in Decision-Making Processes. Current topics in behavioral neurosciences 27: 313–33.

Redish AD. 1999. Beyond the Cognitive Map: From Place Cells to Episodic Memory, MIT Press.

Redish AD. 2004. Addiction as a computational process gone awry. Science 306:1944-1947.

Redish AD. 2013. The mind within the brain: How we make decisions and how those decisions go wrong. Oxford University Press.

Redish AD. 2016. Vicarious trial and error. Nature reviews Neuroscience 17: 147–59.

Regier P, Amemiya S, Redish A. 2015. Hippocampus and subregions of the dorsal striatum respond differently to a behavioral strategy change on a spatial navigation task. Journal of Neurophysiology 114: 1399–1416.

Rescorla R. 2001. Retraining of extinguished Pavlovian stimuli. Journal of Experimental Psychology: Animal Behavior Processes 27(2):115.

Resendez S, Jennings J, Ung R. 2016. Visualization of cortical, subcortical and deep brain neural circuit dynamics during naturalistic mammalian behavior with head-mounted microscopes and chronically implanted lenses. Nature Protocols. 11(3): 566-97

Resulaj A, Kiani R, Wolpert D, Shadlen M. 2009. Changes of mind in decision-making. Nature 461: 263–6.

Reyna V, Ellis S. 1994. Fuzzy-Trace Theory and Framing Effects in Children's Risky Decision Making. Psychological Science 5: 275–279.

Reyna V, Farley F. 2006. Risk and Rationality in Adolescent Decision Making Implications for Theory, Practice, and Public Policy. Psychological Science in the Public Interest 7: 1–44.

Robinson T, Berridge K. 2003. Addiction. Annual review of psychology 54: 25–53.

Robinson TE, Flagel SB. 2009. Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. Biological psychiatry 65(10): 869-873.

Robinson, Berridge. 1993. The neural basis of drug craving: an incentive-sensitization theory of addiction. Brain research Brain research reviews 18: 247–91.

Robinson T.E., Kolb B. 2004. Structural plasticity associated with exposure to drugs of abuse. Neuropharmacology 47, 33-46.

Roese N, Summerville A. 2005. What we regret most... and why. Pers Soc Psychol Bull 31: 1273–85.

Roffler-Tarlov S, Graybiel AM. 1984. Weaver mutation has differential effects on the dopamine-containing innervation of the limbic and nonlimbic striatum. Nature 307: 62-66.

Rozin, P, Schull J. 1988. The adaptive-evolutionary point of view in experimental psychology. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), Stevens' handbook of experimental psychology: Perception and motivation; Learning and cognition (pp. 503-546). Oxford, England: John Wiley & Sons.

Russo S.J., Dietz, D.M., Dumitriu, D., Malenka, R.C., Nestler, E.J. 2011. The addicted synapse: Mechanisms of synaptic and structural plasticity in nucleus accumbens. Trends Neurosci. 33(6) 267-276.

Saint-Cyr JA, Taylor AE, Lang AE. 1988. Procedural learning and neostriatal dysfunction in man. Brain 111(4): 941–959.

Salamone JD, Correa M, Yang J, Rotolo R, Presby R. 2018. Dopamine, Effort-Based Choice, and Behavioral Economics: Basic and Translational Research. Frontiers in Behavioral Neuroscience 12: 52.

Sanfey A, Loewenstein G, McClure S, Cohen J. 2006. Neuroeconomics: cross-currents in research on decision-making. Trends in cognitive sciences 10: 108–16.

Santini E, Ge H, Ren K, Ortiz S de, Quirk G. 2004. Consolidation of fear extinction requires protein synthesis in the medial prefrontal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 24: 5704–10.

Schacter DL, Tulving E. 1994. (eds) Memory Systems (MIT Press, Cambridge, Massachusetts).Schelp SA, Pultorak KJ, Rakowski DR, Gomez DM, Krzystyniak G, Das R, Oleson EB. 2017. A transient dopamine signal encodes subjective value and causally influences demand in an economic context. PNAS 114:E11303–E11312.

Schelp S, Pultorak K, Rakowski D, Gomez D, Krzystyniak G, Das R, Oleson E. 2017. A transient dopamine signal encodes subjective value and causally influences demand in an economic context. Proceedings of the National Academy of Sciences of the United States of America 114: E11303–E11312.

Schmidt B, Papale A, Redish AD, Markus EJ. 2013. Conflict between place and response navigation strategies: Effects on vicarious trial and error (VTE) behaviors. Learning and Memory 20: 130-138.

Schmidt L, Skvortsova V, Kullen C, Weber B, Plassmann H. 2017. How context alters value: The brain's valuation and affective regulation system link price cues to experienced taste pleasantness. Scientific Reports 7: 8098.

Schultz W. 2017. Reward prediction error. Current biology : CB 27: R369–R371.

Shaham Y, Shalev U, Lu L, de Wit H, Stewart J. 2003. The reinstatement model of drug relapse: history, methodology and major findings. Psychopharmacology 168(1-2): 3-20.

Shalev U, Highfield D, Yap J, Shaham Y. 2000. Stress and relapse to drug seeking in rats: studies on the generality of the effect. Psychopharmacology 150(3): 337-346.

Sherry DF, Schacter DL. The evolution of multiple memory systems. Psychology Review 94(4): 439-454.

Shizgal. 1997. Neural basis of utility estimation. Current opinion in neurobiology 7: 198–208.

Sierra-Mercado D, Padilla-Coreano N, Quirk G. 2010. Dissociable Roles of Prelimbic and Infralimbic Cortices, Ventral Hippocampus, and Basolateral Amygdala in the Expression and Extinction of Conditioned Fear. Neuropsychopharmacology 36

Singer R, Zentall T. 2011. Preference for the outcome that follows a relatively aversive event: Contrast or delay reduction? Learning and Motivation 42: 255–271.

Smith JM. 1982. Evolution and the Theory of Games. Cambridge University Press

Smith KS, Graybiel AM. 2013. A Dual Operator View of Habitual Behavior Reflecting Cortical and Striatal Dynamics. Neuron 79: 361–374.

Smith KS, Graybiel AM. 2014. Investigating habits: strategies, technologies and models. Front Behav Neurosci 8: 39.

Smith KS, Graybiel AM. 2016. Habit formation. Dialogues Clin Neurosci 18: 33–43.

Sommer T, Peters J, Gläscher J, Büchel C. 2009. Structure–function relationships in the processing of regret in the orbitofrontal cortex. Brain Structure and Function 213: 535–551.

Squire LR, Knowlton B, Musen G. 1993. The structure and organization of memory. Annual review of psychology 44(1), 453-495.

Stafford D, LeSage M, Glowa J. 1998. Progressive-ratio schedules of drug delivery in the analysis of drug self-administration: a review. Psychopharmacology. 139(3): 169-84

Steiner A, Redish AD. 2012. The road not taken: neural correlates of decision making in orbitofrontal cortex. Frontiers in neuroscience 6: 131.

Steiner A, Redish AD. 2014. Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. Nature Neuroscience 17

Stephens D, Kerr B, Fernández-Juricic E. 2004. Impulsiveness without discounting: the ecological rationality hypothesis. Proceedings Biological sciences 271: 2459–65.

Stephens D, Krebs J. 1986. Foraging Theory. Princeton University Press, Princeton 91.

Stott J, Redish AD. 2014. A functional difference in information processing between orbitofrontal cortex and ventral striatum during decision-making behaviour. Philosophical transactions of the Royal Society of London Series B, Biological sciences 369.

Strough J, Bruin W, Parker A, Karns T, Lemaster P, Pichayayothin N, Delaney R, Stoiko R. 2016. What were they thinking? Reducing sunk-cost bias in a life-span sample. Psychology and Aging 31: 724.

Strough J, Karns T, Schlosnagle L. 2011a. Decision-making heuristics and biases across the life span. Annals of the New York Academy of Sciences 1235: 57–74.

Strough J, Mehta C, McFall J, Schuller K. 2008. Are Older Adults Less Subject to the Sunk-Cost Fallacy Than Younger Adults? Psychological Science 19: 650–652.

Strough J, Schlosnagle L, DiDonato L. 2011b. Understanding Decisions About Sunk Costs From Older and Younger Adults' Perspectives. The Journals of Gerontology: Series B 66B: 681–686.

Strough J, Schlosnagle L, Karns T, Lemaster P, Pichayayothin N. 2014. No Time to Waste: Restricting Life-Span Temporal Horizons Decreases the Sunk-Cost Fallacy. Journal of Behavioral Decision Making 27: 78–94.

Suri R, Schultz W. 1999. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience 91: 871–90.

Sutton R, Barto A. 1998. Reinforcement learning: An introduction. MIT Press, Cambridge, MA.

Suzuki A, Josselyn S, Frankland P, Masushige S, Silva A, Kida S. 2004. Memory Reconsolidation and Extinction Have Distinct Temporal and Biochemical Signatures. J Neurosci 24: 4787–4795.

Sweis BM, Abram SV, Schmidt BJ, Seeland KD, MacDonald AW, Thomas MJ, Redish AD. 2018c. Sensitivity to "sunk costs" in mice, rats, and humans. Science 361: 178-181.

Sweis BM, Larson EB, Redish AD, Thomas MJ. 2018e. Altering gain of the infralimbic to accumbens shell circuit alters economically dissociable decision-making algorithms. PNAS 201803084.

Sweis BM, Redish AD, Thomas MJ. 2018d. Prolonged abstinence from cocaine or morphine disrupts separable computation-specific valuations. Nature Communications 9, 2521.

Sweis BM, Thomas MJ, Redish AD. 2018a From memory to decision-making: Addiction as a heterogeneous disease of computation-specific valuation processes. Learning & Memory (in press).

Sweis BM, Thomas MJ, Redish AD. 2018b. Mice learn to avoid regret. PLoS Biology 16(6): e2005853.

Talmi D, Seymour B, Dayan P, Dolan RJ. 2008. Human Pavlovian–instrumental transfer. Journal of Neuroscience, 28(2): 360-368.

Thaler R. 1999. Mental accounting matters. J Behav Dec Making. 12: 183-206

Thomas MJ, Beurrier C, Bonci A, Malenka RC. Long-term depression in the nucleus accumbens: a neural correlate of behavioral sensitization to cocaine. Nature Neuroscience 4(12): 1217-23.

Thomas MJ, Kalivas PW, Shaham Y. 2008. Neuroplasticity in the mesolimbic dopamine system and cocaine addiction. Br J Pharmacol 154: 327–42.

Thomas MJ, Malenka RC, Bonci A. 2001. Modulation of Long-Term Depression by Dopamine in the Mesolimbic System. J Neuroscience 20(15): 5581-5586.

Thomas MJ, Malenka RC. 2000. Modulation of long-term depression by dopamine in the mesolimbic system. Philos Trans R Soc Lond B Biol Sci. 358(1432): 815-819

Tiffany ST. 1995. Potential functions of classical conditioning in drug addiction.

Tolman E. 1939. Prediction of vicarious trial and error by means of the schematic sowbug. Psychological Review 46: 318–336.

Tolman EC. 1948. Cognitive maps in rats and men. Psychological Review 55: 189–208.

Torges C, Stewart A, Nolen-Hoeksema S. 2008. Regret resolution, aging, and adapting to loss. Psychology and Aging 23: 169.

Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RGM. 2007. Schemas and memory consolidation. Science 316: 76-82.

Tye KM. 2012. Glutamate inputs to the nucleus accumbens: does source matter? Neuron. 76(4): 671-73

Tzschentke TM. 2007. Review on CPP: Measuring reward with the conditioned place preference (CPP) paradigm: update of the last decade. Addiction Biology. 12(3-4): 227-462

van der Meer MAA, Johnson A, Schmitzer-Torbert NC, Redish AD. 2010. Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. Neuron 67: 25-32.

van der Meer MAA, Kurth-Nelson Z, Redish AD. 2012. Information processing in decision-making systems. The Neuroscientist 18(4): 342-359.

van der Meer MAA, Redish AD. 2009a. Covert Expectation-of-Reward in Rat Ventral Striatum at Decision Points. Frontiers in integrative neuroscience 3: 1.

van der Meer MAA, Redish AD. 2009b. Low and high gamma oscillations in rat ventral striatum have distinct relationships to behavior, reward, and spiking activity on a learned spatial decision task. Front Integr Neurosci. 3:9

van der Meer MAA, Redish AD. 2011. Theta phase precession in rat ventral striatum links place and reward information. The Journal of neuroscience : the official journal of the Society for Neuroscience 31: 2843–54.

van Wingerden M, Marx C, Kalenscher T. 2015. Budget Constraints Affect Male Rats' Choices between Differently Priced Commodities. PLoS One 10(6): e0129581.

Vandaele Y, Cantin L, Serre F, Vouillac-Mendoza C, Ahmed SH. 2016. Choosing under the influence: a drug-specific mechanism by which the setting controls drug choices in rats. Neuropsychopharmacology. 42(2): 646-657

Volkow ND. 2012. National Institute on drug abuse. Corsini Encyclopedia of Psychology.

Walters C, Redish AD. 2018. Chapter 8. "A Case Study in Computational Psychiatry Addiction as Failure Modes of the Decision-Making System." in Comp Psych: Mathematical modeling of mental illness, (A. Anticevic and J. Murrary, eds). Elsevier

Wang JX, Cohen NJ, Voss JL. 2015. Covert rapid action-memory simulation (CRAMS): A hypothesis of hippocampal--prefrontal interactions for adaptive behavior. Neurobiol Learn Mem 117: 22-33.

Wassum KM, Izquierdo A. 2015. The basolateral amygdala in reward learning and addiction. Neurosci Biobehav Rev 57: 271–83.

Weatherhead P. 1979. Do Savannah Sparrows Commit the Concorde Fallacy? Behavioral Ecology and Sociobiology 5: 373–381.

Weinshenker D, Schroeder JP. 2007. There and back again: A tale of norepinephrine and drug addiction. Neuropsychopharmacology 32: 1433-1451.

Weller J, Levin I, Rose J, Bossard E. 2012. Assessment of Decision-making Competence in Preadolescence. Journal of Behavioral Decision Making 25: 414–426.

Wikenheiser AM, Stephens DW, Redish AD. 2013. Subjective costs drive overly-patient foraging strategies in rats on an intertemporal foraging task. PNAS 110(20): 8308-8313.

Wilson T, Gilbert D. 2005. Affective forecasting: Knowing what to want. Current Directions in Psychological Science. 14(3): 131-134

Winkler D. 1991. Parental investment decision rules in tree swallows: parental defense, abandonment, and the so-called Concorde Fallacy. Behavioral Ecology 2: 133–142.

Wolf ME. 2016. Synaptic mechanisms underlying persistent cocaine craving. Nat Rev Neurosci 17: 351–65.

Yin HH, Knowltown BJ, Balleine BW. 2004 Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. European Journal of Neuroscience 19(1): 181-189.

Yizhar O, Adamantidis A. 2018. Cell Type-Specific Targeting Strategies for Optogenetics. Optogenetics: A Roadmap. Neuromethods

Zapata A, Minney VL, Shippenberg TS. 2011. Shift from goal-directed to habitual cocaine seeking after prolonged experience in rats. J Neurosci 30(46): 15457-15463.

Zeeb F, Baarendse P, Vanderschuren L, Winstanley C. 2015. Inactivation of the prelimbic or infralimbic cortex impairs decision-making in the rat gambling task. Psychopharmacology 232: 4481–91.

Zeelenberg M, Pieters R. 2007. A theory of regret regulation 1.0. J Consumer Psych 17: 3–18.

Zweifel L, Argilli E, Bonci A, Palmiter RD. 2008. Role of NMDA receptors in dopamine neurons for plasticity and addictive behaviors. Neuron. 59(3): 486-96