Dopaminergic control of neuroeconomic decision making

A DISSERTATION SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Adrina Kocharian

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisors: Patrick E. Rothwell, PhD & A. David Redish, PhD

Acknowledgments

I want to express deep gratitude to several individuals who have offered me strength and support throughout my training. This would not have been possible without you.

My first ever mentor, Kurt Kroeger taught me scientific professionalism while I was a high schooler interning in a lab with no experience. To my mentors in university, Dr. Nori Kasahara and Dr. Jon Voutila, thank you for giving me independence to develop my interests in neuroscience.

I want to especially thank Dr. Chris Evans who taught me about the neurobiology of drugs of abuse and led an outreach course educating high schoolers about how drugs interact with the brain. Many of my research interests developed through these courses and propelled me toward this career path.

I experienced some of my most profound personal and professional growth during my postbac at the National Institutes of Health. I greatly thank Dr. Andrew Holmes for striking a delicate balance between challenging me scientifically and indulging in Kardashian drama. Thank you for setting high expectations and giving me my start. Dr. Anna Radke, thank you for being such an exemplary role model for me and introducing me to both the world of dopamine and the University of Minnesota's Graduate Program in Neuroscience. Dr. Lindsay Halladay, thank you for the opportunity to work on such an interesting project with you. Thank you to my Holmes lab homies as well: Dr. Katie Kaugers, Dr. Shana Bug Silverstein, Dr. Ayesha Sengupta, the Abbes (Liberman and Postle), Aaron, and Emma.

I will forever be appreciative to Dr. Yoji Shimizu, who believed in my abilities enough accept me into the Medical Scientist Training Program. The program is so fortunate to have you lead us. To

the entire MSTP leadership, thank you for guiding us through this intensive training. Thank you especially to Nick Berg, Susan Shurson, Dr. Linda McLoon, Dr. Bryc Binstadt, and Dr. Suma Jacob. To my MSTP cohort (Roberto, Bayardo, Natalie, Fathima, Minna, Liz, and Jenna), I am so grateful for our little family.

Thank you to my various mentors and support system during my early clinical years of medical school. Dr. Alik Widge, Dr. Sofia Lyford-Pike, Dr. Rebecca Speltz, thank you for teaching me how to approach clinical problems. And Dr. Widge, thank you for always encouraging me to maintain my interests in policy work and advocacy.

To my trauma-bonded friend group from medical school— I could not be prouder of you all making it to the other side. I would trust each of you (well, most of you) with my life. Luke, Tamara, Beret, Elle, Noey, Siddhant, Marin, Ally, Hannie, James, Dylie, I hope we stay in each other's lives forever. Please be nice to me when you are my attendings.

To my partner Gabe—thank you for being by my side throughout this (very long) journey. I often look back at our early days together when we wondered in angst about what graduate programs we would be pursuing. I cannot imagine getting through the hard parts without you. I have learned a lot about myself through you and your mind is an extension of my own.

I want to thank the Graduate Program in Neuroscience for the opportunity to train with under such inspiring scientists and alongside such impressive trainees. Dr. Lucy Vulchanova, thank you for being such a support system and encouraging me to pursue some unconventional paths in my training. Thank you to Rebecca, John, and Lateeph for keeping our program afloat and being a constant resource for me. Thank you to Dr. Erin Larson for your technical support and wisdom.

To my graduate school cohort, I treasure our memories at Itasca and beyond. Specifically, thank you to Jackson, Roberto, and Lizzie. Lizzie, I cannot imagine this experience without you or the Real Housewives of Salt Lake City.

A tremendous thanks to the members of the Rothwell lab who have commiserated and celebrated with me at various points throughout graduate school. To some of the OGs, Dieter, Cassie, Brian, Emilia, Bailey, Carlee, and Marc, your warmth and thrill for science were what compelled me to join the Rothwell lab. I am so lucky to have been able to learn from you. To the newer crew, Filip, Ross, Rachel, Nicole, Anisha, Mohammed, Elysia, I am so happy to see the lab continue to grow with enthusiastic, good-hearted people. I will forever be indebted to the several undergraduates who helped me complete this work. Jacqueline and Hailie, especially, I could not have made it through these time-intensive experiments without your help. It has been an absolute pleasure working with you, teaching you, and learning from you both. I look forward to seeing what lies ahead in your careers.

I want to thank Dr. Brian Sweis for all of the guidance when I was getting my project started. Thanks for talking me down when I was not sure about taking on this work.

To my thesis committee, Dr. Mark Thomas, Dr. Alik Widge, and Dr. Nicola Grissom, I greatly appreciate you being an advocate for me. Thank you for helping to steer this project toward completion. Mark, I especially thank you for allowing me to use your equipment throughout my thesis work.

Finally, to my advisers, Dr. Patrick Rothwell and Dr. David Redish— thank you both for taking on an excited graduate student with an idea. Your expertise and approach to mentorship complemented one another, and I am proud of the work we did as a team.

Patrick— I have such profound respect for your commitment to supporting students. You have taught me that it is possible to be both nurturing and productive. I admire your intuitions for a project, and your ability to see ten steps ahead. Thank you for always being so assuring at times when I doubted the process. And thank you for always leading by example.

Dave— I truly value your visionary approach to science. You have taught me how to think deeply and curiously. Thank you for pushing me and letting me push back. You are tough and have made me tougher. Thank you for always finding time to talk science and being available at a moment's notice in a science "crisis". Thank you for always advocating for me, and for genuinely feeling pride in my successes.

Dedication

I would like to dedicate this thesis to my parents and brothers.

To my parents—I am constantly in awe of you for leaving your lives as scientists behind in Armenia and coming to the states to start all over. Science is the family business. You pushed me and believed in my abilities, and I will never be able to express my full appreciation to you both.

To my brothers—I love you both. We are cosmically bonded in ways that cannot be described in words. You mean the world to me.

This thesis is for all of us. It is a lesson to my past self that I can do hard things, and a reminder to my future self to keep chasing questions.

Abstract

Dopamine in the nucleus accumbens is an important neural substrate for valuation and decision-making. Dominant theories generally discretize and homogenize decision-making, when it is in fact a continuous process, with evaluation and re-evaluation components that extend beyond simple outcome prediction into consideration of past and future value. Furthermore, individual animals use distinct strategies to achieve their goals, requiring different computational processes. Extensive work has examined mesolimbic dopamine in the context of reward prediction error, but major gaps persist in our understanding of how dopamine tracks imagined past and future rewards to influence decision confidence. Moreover, there is little consideration of strategy-dependent differences in value processing that may shape dopaminergic encoding. In the studies presented in this dissertation we used an economic foraging task in mice, and found that strategy-specific dopamine dynamics reflected decision confidence during evaluation, as well as both past and future counterfactual value during re-evaluation. We found that inhibition of dopamine terminals altered counterfactual processing during re-evaluation. Individually-tailored optogenetic stimulation of mesolimbic dopamine terminals altered decision confidence during evaluation and carried over to counterfactual re-evaluation, in a strategy-specific manner. We provide evidence that mesolimbic dopamine is tightly linked to decision confidence and counterfactual information, through signals that go beyond reward prediction errors to more complex encoding of imagined past and future value.

Table of Contents

List of Figures

Chapter 2: Individual differences in decision strategy shape how mesolimbic dopamine encodes counterfactual future to influence choice confidence

List of Tables

List of Abbreviations

NAc – nucleus accumbens

VTA–ventral tegmental area

SNc – substantia nigra pars compacta

D1-MSN – medium spiny neuron expressing dopamine type-1 receptor

D2-MSN – medium spiny neuron expressing dopamine type-2 receptor

US – unconditioned stimulus

CS–conditioned stimulus

OZDM – Offer Zone Decision Maker

WZDM – Wait Zone Decision Maker

ACTR – Adaptive Rate Cost of Performance to Reinforce

ANCCR – Adjusted Net Contingency for Causal Relations

KCS – Kutlu-Calipari-Schmajuk

RPE – reward prediction error

TDRL – temporal difference reinforcement learning

VTE – vicarious trial and error

Chapter One

Introduction

1.1 Counterfactual reasoning and neuroeconomic decision-making

Neuroeconomics

The decision-making landscape can be appreciated through the lens of neuroeconomics. The field of neuroeconomics is multidisciplinary and aims to describe the computational mechanisms by which the brain encodes and processes value (Loewenstein, Rick, & Cohen, 2008)(Camerer, Loewenstein, & Prelec, 2005) . The goal in doing so is to elucidate how individuals process and utilize value to ultimately guide decision-making. To understand the neural underpinnings of decision computations is to understand how agents may process and act on decisions, especially when encountering conflicting options. Leveraging experimental methods spanning economics, psychology, neuroscience, and computer science, neuroeconomics seeks to identify the neurobiological and computational variables used to make different decisions (Rangel & Clithero, 2013).

Multiple computational systems within the brain can process information in parallel. Distinct decisional computations may ultimately yield seemingly identical decision outcomes, emphasizing the importance of studying the means and mechanisms by which value is derived. Separable value

systems may map onto discrete neural circuits or neurochemical representations in the brain. Therefore, it is critical to explore not only *what* decisions are made, but also *how* decisions are made.

Counterfactual Reasoning

Understanding the causality of an agent's actions is crucial for learning about the environment and executing decisions. Predictive knowledge about the future can be exploited to guide decision-making in two major ways: deciphering the relationship between chosen actions and outcomes and deciphering the relationship between alternative actions and the outcomes they *could* have produced had they been chosen. The latter is a form of causal inference called counterfactual reasoning which is a major feature of decision-making. Through counterfactual thinking, one compares "what has occurred" with "what might have occurred". This can often result in experiences of regret (when an outcome is much worse than what could have been), or relief (when an outcome is much better than what could have been).

Importantly, counterfactual experiences like regret require the acknowledgement of agency in the outcome at hand. That is, experiencing a worse than expected outcome over which we had no control merely yields disappointment, whereas experiencing an outcome that is worse than what could have been had we made a different decision may yield regret. Therefore, regret requires a sense of personal responsibility.

Imagine the following scenario: You are a student who is constantly faced with balancing a full course load, several extracurriculars, and social life. In two days, you have a chemistry exam, but all your friends are going to a party tonight. You hesitate, deliberate, but ultimately decide to join your friends. The next night you are stuck cramming for your chemistry exam. After a full night of studying, you take your exam and receive a C+. In this case, the actual information you received is the grade of a C+. However, you may contextualize this outcome within a counterfactual framework of what could have happened had you studied more and think, "This is awful. If had missed the party and studied more, I would have received an A." Or perhaps, you may contextualize this outcome in a counterfactual framework of what could have happened had you studied less and think, "Thank goodness. If I hadn't crammed, I may have failed." Counterfactual thoughts underlie experiences of regret and relief, which are salient determinants of decision-making.

Psychology of regret

From imagining the paths untraveled in our careers to deducing the alternate probability of having caught the flight that you ended up missing, counterfactual thinking is pervasive in our daily lives. Counterfactual thoughts emerge early in development in children as young as two years old who learn to consider simple alternate outcomes like, "What if mommy had taken her shoes off, would the floor be dirty?" (T. P. German & Nichols, 2003) They form the foundation for creativity and fantasy as well as high order cognitive functions like exploring counterexamples to deductions (Byrne, Espino, & Santamaria, 1999).

Interestingly, in humans, regret-inducing counterfactuals (e.g. “I should have brought my umbrella, now I’m drenched!”) are much more common than relief-inducing counterfactuals (e.g. “Thank goodness I brought my umbrella, now I’m dry!”) (Roese, Sanna, & Galinsky, 2012). Perhaps this is because the perceived saliency of being drenched in a rainstorm greatly outweighs the perceived saliency of staying dry. We know from decades of psychology research that negative affect produces stronger effects than positive affect. For example, having a bad day has a longer lasting impact on mood than having a good day (Sheldon, Ryan, & Reis, 1996). On a less subjective measure, unpleasant odors elicit greater facial expressions than pleasant one(Gilbert, Fridlund, & Sabini, 1987). And of course, a single traumatic experience can impart severe and lasting effects on a person’s psychological health even if followed by a lifetime of positive experiences (Brickman, Coates, & Janoff-Bulman, 1978).

However, regret is even more striking than any of these examples because of the recognition of agency in one’s own misfortune or demise. The fact that you could have simply left the house 15 minutes earlier and made it on time for your flight yields greater negative affect than you being notified your flight was canceled. Both scenarios involve you missing your flight, yet the former induces more significant frustration.

It is no surprise, then, that humans develop regret aversion. This aversion can alter decision making in individuals by leading them to 1) re-evaluate an outcome and attempt to *undo the choice*, and 2) *avoid the choices* that are anticipated to lead to regret later. That is, the possibility of committing to a failure or missing out on a better opportunity can repel individuals from making certain decisions in order to minimize regret.

Undoing the choice that caused regret

Attempting to undo choices relates to whether the regret experienced is due to action or inaction. Regrets due to personal action may involve thinking along the lines of "I should not have bought that $500 dress", whereas regrets due to personal inaction may involve thoughts like "I should have ordered something healthier". Both scenarios induce regret that may propel the decision-maker to rectify their mistakes (i.e., return the dress or eat only half of the burger). However most people find that the pain of regrettable action is more salient and proximal than the pain of regrettable inaction, and are therefore more likely to act to correct regrets due to action (Gilovich & Medvec, 1995).

Avoiding the choice that may cause regret

After making regretful choices, humans often adjust their future decision-making to prevent the unsatisfying experience of future regret. Avoidance of anticipatory regret is a decision-making strategy that can be very adaptive and motivational. For instance, the anticipation of missing out on a workout, "if I don't go to my workout class, I will regret it later", is a motivator for engaging in healthy behavior. In fact, greater anticipatory regret predicts stronger intention to exercise (Abraham & Sheeran, 2004).

Neuroscience of regret

While the impacts of counterfactual reasoning on decision-making are clear, there is still much to appreciate about how these computational processes are encoded in the brain. In order for an agent to engage in counterfactual reasoning, they must invoke an imagination or representation of an alternative condition. These alternate states carry value that are weighed against selected states. If the selected state is worse than what could have been in a counterfactual alternate condition, experiences of regret may follow. Understanding the neurobiological computations of experiences like regret may help elucidate circuit or neurochemical mechanisms underlying psychiatric diseases in which decision-making systems break down.

We know, for example, that patients who have experienced focal lesions to the prefrontal cortex experience difficulty organizing problems spaces, shifting between different mental sets, and planning in a structured manner compared to healthy controls (Goel, Grafman, Tajik, Gana, & Danto, 1997). Similarly, patients with ventromedial prefrontal cortex lesions have difficulty updating their decision strategies by learning from past mistakes on the Iowa Gambling Task (Bechara, Damasio, Damasio, & Anderson, 1994).

One such example of a gambling task used to probe counterfactual reasoning gives participants the choice of selecting a risky choice for a possible higher monetary reward or a safer choice for a probable lower monetary reward. Using this task, Camille et al. (2004) showed that the experience of gains or losses depends on the alternative outcomes. That is, when healthy participants received feedback only on their selected choice outcome, they experienced disappointment; however, when

they received complete feedback on the alternative choice's outcome, they experienced regret and stronger physiological responses, as measured by skin conductance. Interestingly, patients with lesions to the orbitofrontal cortex did not show this heightened emotional response to complete feedback, nor did they exhibit physiologic markers of emotional arousal (Camille et al., 2004).

Just as focal lesions inform our understanding of counterfactual processing in the brain, so too do neurological and psychiatric disorders. Patients with schizophrenia, much like patients with lesioned orbitofrontal cortex, experience less regret and less anticipation of negative consequences of their choices than healthy controls on a regret gambling task (Larquet, Coricelli, Opolczynski, & Thibaut, 2010). While Parkinson's and Huntington's disease are most noted for their motoric symptomatology, the cognitive dysfunctions that ensue with disease progression can greatly impair decision-making. Patients with Parkinson's or Huntington's disease generate fewer counterfactual thoughts of "if only" or "what if", and struggle with inferring counterfactual information than healthy controls (McNamara, Durso, Brown, & Lynch, 2003)(Solca et al., 2015).

Some studies have unveiled possible circuits and neurochemicals implicated in counterfactual reasoning. For example, using functional magnetic resonance imaging (fMRI) on humans performing a risky gambling task, Coricelli and others discovered that regret-inducing outcomes correlated with increased activity in the medial orbitofrontal cortex, anterior cingulate, and hippocampus. Moreover, as participants gained familiarity with the task, they became regret averse, which also correlated with medial orbitofrontal cortex activation. This suggests that some of the same brain regions may contribute both to the anticipation/avoidance of regret as well as the actual experience of regret itself (Coricelli et al., 2005).

While fMRI illuminates brain regions that may be involved in counterfactual thinking, there is added value in understanding the neurochemical correlates of these processes. One critical candidate is dopamine, whose actions are broadly implicated in decision-making (which will be discussed in greater detail later).

Some groups have even recorded dopamine levels in patients with Parkinson's disease who have existing deep brain stimulating electrodes in their striatum, while they play a sequential investment game. During the task, patients make various investment decisions by betting 0-100% of their portfolio, after which they experience possible returns or losses proportional to the change in market price. Using fast-scan cyclic voltammetry, Kishida et al. were able to show that dopamine fluctuations integrate expectation violations with counterfactual violations of what could have been had they made better or worse investments (Kishida et al., 2016). The same approach of using human voltammetry was applied to studying essential tremor in Parkinson's disease patients with deep-brain stimulation electrodes in the caudate with and without a history of alcohol use disorder. On a "Sure bet or gamble" task, participants chose between a guaranteed option or a 50% gamble on each trial while dopamine fluctuations were monitored. Dopamine to counterfactual information about alternate options on gamble trials differed between patients with and without a history of alcohol use disorder (Liebenow et al., 2022).

Since dopamine may be an important player in counterfactual thinking, one group attempted to examine the effects of the dopamine partial agonist, aripiprazole, on alternative outcome learning. Patients with Tourette syndrome treated and untreated with aripiprazole performed a probabilistic learning task in which they had to learn which stimuli resulted in a correct choice that is more

often rewarding than punishing. After making a choice, patients received either partial feedback about the outcome of their chosen option, or complete feedback about the outcome of both the chosen and unchosen outcomes. Complete feedback produced counterfactual thoughts of what alternate option could have been selected. Interestingly patients treated with aripiprazole did not improve their learning after receiving complete counterfactual feedback compared to untreated patients and healthy untreated controls (Salvador et al., 2017).

Operationalizing counterfactuals in non-humans

While these human studies form the foundation for understanding the neural underpinnings of counterfactual thought, there are several caveats we must consider. First, the studies that probe neural activity of neurochemical levels in humans are conducted in patients who have existing disease. Often, as in the cases above, the patients have a deep-brain stimulation device implanted in a region precisely to correct an existing neural dysfunction (i.e., neurodegeneration of dopamine neurons), so the measurements taken from these individuals must be considered in the context of the existing disease. Second, there are very few opportunities to causally manipulate brain activity or neurotransmission in humans for ethical and practical reasons.

These limitations pose significant barriers to understanding the contributory mechanisms of counterfactual reasoning. Thus, operationalizing computational processes using non-human animals in controlled laboratory settings is critical to addressing these types of questions. Beyond the methodological advantages to using non-human animals, these approaches can elucidate the evolutionary roots of decision-making (Kalenscher & van Wingerden, 2011).

To this end, various decision-making tasks have been adapted for use in non-human primates and rodents to probe counterfactual reasoning across species. In a computerized probabilistic rock-paper-scissors game, rhesus monkeys learn to optimize their behavior by adjusting their choices based on both the reward probability of their choice (say, "paper" when the opponent computer selected "scissors"), but also based on the reward probability of the unchosen outcome (in this case "rock" which would have beaten the computer, and thus resulted in greater reward delivery). The experimenters found that both dorsolateral prefrontal cortex and orbitofrontal cortex single-unit activity encoded signals related to both choices and unchosen alternatives. Furthermore, they found that neuronal activity of dorsolateral prefrontal cortex changed its activity significantly scaling with the hypothetical winning payoffs (Abe & Lee, 2011).

To further explore the computations executed by cortico-subcortical circuits to integrate actual and counterfactual values, another group recorded single-unit activity from the orbitofrontal cortex, ventral striatum, and midbrain dopamine neurons. Monkeys performed a two-step economic decision-making task in which they had to choose or not choose an option based on its value. After choosing or not choosing in the first step, the second option (with better or worse associated reward) was presented. If the monkey had chosen the first option, it would not be permitted to choose the second option, and conversely, if the monkey had not chosen the first option it would be obligated to choose the second option. Therefore, the chosen option carries actual information while the first or second unchosen option carries counterfactual information. They found that orbitofrontal cortex neurons represented actual and counterfactual values in an opposite and antagonistic manner (Yun et al., 2023).

Fewer studies have explored counterfactual processing in rodents. The first evidence of regret-like behavior in rodents was reported by Steiner and Redish in 2014. In this study, rats foraged for food at various costs (delays) on a spatial economic task (Restaurant Row) wherein rats could serially make the decision to accept or reject delays for flavored pellets. If rats encountered high-cost delays after previously rejecting a low-cost option, they would look backwards at their missed opportunity (although the previous offer was now unavailable), and recover time by rushing to eat food on the high-cost trial. When rats oriented backwards toward the missed low-cost opportunity, neurons in the orbitofrontal cortex and ventral striatum represented the previous low-cost zone.

Years later, Sweis et al. adapted this task for use in mice. An important modification to this task was an addition of initial accept/reject Offer zone decision preceding an earn/quit Wait zone decision. In this version, after accepting an offer delay in the Offer Zone, mice had to wait for the countdown to reward in the Wait zone. This allowed for a two-step decision-making process whereby mice could accept offers but demonstrate change-of-mind behaviors of quitting mid-countdown. This regret-like process of quitting was followed by a propensity to make up for lost time by altering willingness to wait on subsequent trials and rushing to consume food. As training progressed, mice learned to avoid "regret-inducing" experiences by developing future-planning strategies that would avoid wasting time accepting offers that they would later quit (Sweis, Thomas, & Redish, 2018).

Another study examining counterfactual reasoning in rodents probed learning about action-outcome relationships. Training rats on a modified version of the Pavlovian-Instrumental Transfer

test, experimenters conducted tests of negative predictions, over-expectation, and backwards conditioning to probe the effects of excitatory and inhibitory cues that predicted whether an action would or would not result in an outcome, respectively. Excitatory cues resulted in increased responding to the lever to that predicted an outcome, however inhibitory cues resulted in increased responding to the alternate lever that predicted absence of the inhibited outcome (Laurent & Balleine, 2015).

While many investigators have explored decision-making in non-human animals, the contributions of counterfactual information to economic valuation is greatly understudied. In order to fully appreciate the dynamics of decision-making, it is important to characterize not only the actual outcomes of choice but also the counterfactual paths unchosen. Moreover, we must attempt to probe the neural mechanisms that encode this counterfactual information in order to unveil the computational processes that may underly these salient experiences.

1.2 Dopaminergic neurotransmission and circuits

Dopamine mechanisms

As a major catecholaminergic neurotransmitter in the brain, dopamine exerts its effects across various neural circuits. However, dopamine itself is produced by a small subset of neurons within the ventral tegmental area (VTA) and substantia nigra (SNc) of the midbrain.

A derivative of tyrosine, DOPA is converted to dopamine, which is in turn a precursor to norepinephrine. Dopamine gets packaged into vesicles for neurotransmission by vesicular monoamine transporter (VMAT-2) and trafficked to the plasma membrane at release sites. Upon release at terminals, dopamine has the ability to interact with various dopamine receptors subptyes: D1, D2, D3, D4, and D5, all of which are G-protein-coupled. The D1 and D2 receptors are the most prevalent subtypes throughout the brain (Levey et al., 1993)(Cameron & Williams, 1993), highly expressed on medium spiny neurons (MSNs) in the striatum, and differentially mediate the effects of dopamine signaling. These receptors form two dopamine receptor families: the D1-like family is composed of D1 and D5 which are coupled to G_s and G_q, while the D2-like family is composed of D2-D4 which are coupled to G_i and G_o. D1 activation results in increased intracellular cAMP, leading to downstream increases in neuronal excitability, while D2 activation decreases intracellular cAMP and neuronal excitability. These receptors differ in their affinities with D2 receptors having approximately 100-fold greater affinity to dopamine than D1 receptors. It is thus posited that “tonic” baseline levels of dopamine release in the nanomolar range activate D2

receptors, whereas “phasic” micromolar fluctuations in dopamine release activate D1 receptors (C. L. German, Baladi, McFadden, Hanson, & Fleckenstein, 2015).

Extracellular levels of dopamine are regulated by various clearance mechanisms. Passive diffusion, degradation through monoamine oxidase, inactivation through catechol-O-methyltransferase all contribute to the clearance of dopamine from the synapse. However the major contributor to dopamine clearance is the dopamine transporter, DAT. DAT is expressed presynaptically, and shifts between outward- and inward-facing conformations to transport dopamine across the membrane and influence the duration and extent of postsynaptic signaling. Another way dopamine neurons regulate release dynamics is through the presynaptic D2 receptor at axons. As dopamine binds these autoreceptors, dopamine release is negatively regulated through an auto-inhibitory feedback mechanism.

Dopaminergic circuits

Dopaminergic axons project most densely to three major terminal regions: projections to the cortex, ventral striatum, and dorsal striatum and form the mesocortical, mesolimbic, and nigrostriatal pathways, respectively. Interestingly these projections differ in their physiologic and molecular properties, as well as their inputs. VTA and SNc dopamine neurons express differing levels of ion channels and generate distinct firing patterns of activity. Nigrostriatal dopamine neurons arborize more widely than mesolimbic or mesocortical projections. SNc dopamine neurons show enhanced vulnerability to degeneration compared to VTA dopamine neurons

(Matsuda et al., 2009)(Chan, Gertler, & Surmeier, 2010)(Zampese, Galtieri, Schumacker, & Surmeier, 2016).

Beyond distinctions between projections, dopamine neurons *within* a region or projection also exhibit great heterogeneity. The VTA, especially, is comprised of a very complex microcircuitry. Distinct subtypes of dopamine neurons express genes such as Vglut2, Sox6, Aldh1a1 (Poulin et al., 2014)(Tiklová et al., 2019). It has also been shown that some dopamine subpopulations, and not others, co-release other neurotransmitters like GABA or glutamate (Hnasko et al., 2010)(Stuber, Hnasko, Britt, Edwards, & Bonci, 2010). While nearly all VTA dopamine neurons show $GABA_B$- receptor mediated responses (Labouèbe et al., 2007), there is some heterogeneity in electrophysiological properties among dopamine neurons in the VTA. While most possess "classic" dopamine-neuron properties (e.g., long action potential durations and hyperpolarization-activated cation currents), others lack these properties (Lammel et al., 2008).

1.3 Dominant theories on dopamine dynamics during decision-making

While the cortex and several other regions like the amygdala, hippocampus, and cerebellum receive dopaminergic input, the striatum is the most densely innervated and most studied dopaminergic projection. Mesolimbic and nigrostriatal dopamine are known to be critical in regulating movement, motivation, and reinforcement learning.

Movement

The degeneration of SNc dopamine neurons in Parkinson's disease has served as evidence for the important role of dopamine in facilitating spontaneous movement. In fact, formative experiments conducted in the 1950s by Arvid Carlsson and colleagues found that intravenous infusion of the dopamine precursor, 3,4-dihydroxyphenylalinine (DOPA), could attenuate akinesia effects from reserpine. The rescue of movement was correlated specifically with brain dopamine, not norepinephrine. Just a year later, Carlsson and collaborators discovered that the majority of dopamine was located in striatal regions, which contained very little norepinephrine. These findings led to the discovery of dopamine as an independent neurotransmitter (Carlsson, Lindqvist, & Magnusson, 1957)(Carlsson, Lindqvist, Magnusson, & Waldeck, 1958)(Bertler & Rosengren, 1959).

Since then, numerous studies have implicated dopamine in action selection, vigor, movement initiation, and velocity (R. S. Lee, Mattar, Parker, Witten, & Daw, 2019). The majority of dopamine neurons involved in movement are located in the SNc. Essentially, nigrostriatal

dopamine is less critical for the kinematic aspect of movements, rather it is essential for initiation and self-motivated of movement. A classic example of this is that patients with Parkinson's Disease have difficulty initiating walking; however, these same patients can readily take steps on a treadmill, climb stairs, or jump up at the sound of a fire alarm (Glickstein & Stein, 1991).

Motivation

Another appreciated function of dopamine is "ramping" during motivational pursuit. Ramps in dopamine have been observed as animals gain spatial or temporal proximity to rewards. These ramps have predominantly been reported in ventral striatal regions through recordings of extracellular dopamine using microdialysis (Ostlund, Wassum, Murphy, Balleine, & Maidment, 2011)(Sokolowski, Conlan, & Salamone, 1998), fast-scan cyclic voltammetry (Howe, Tierney, Sandberg, Phillips, & Graybiel, 2013)(Collins et al., 2016), and fiber photometry recordings of calcium signals from dopamine axons or extracellular dopamine levels (Mohebi et al., 2019) (Kim et al., 2020).

Motivational pursuit is related to incentive salience, a term describing the psychological process underlying the "wanting" of rewards. Interestingly, dopamine depleted rats show normal "liking" of pleasant tastes, measured by assessing hedonic orofacial expressions, but display diminished "wanting" for the sweetness, measured by reduced motivation to voluntarily consume (Berridge & Robinson, 1998). This type of wanting can be triggered by reward-related cues—many studies have found that increases in NAc dopamine to reward-predicting cues is associated with motivation during reward-seeking behavior. For instance, NAc dopamine measurements taken

from rats lever-pressing to receive food can vary based on satiety level. That is, animals show greater increases in dopamine when they are in a hungry, higher motivational state than when they are sated in a lower motivation state, to the same exact reward (Ostlund et al., 2011). Additionally, presentation of a drug- or reward-paired cue can invigorate an animal to increase self-administration, and this process is dependent on dopamine (Wyvell & Berridge, 2000)(Ostlund & Maidment, 2012)(Wassum, Ostlund, Loewinger, & Maidment, 2013a).

Reinforcement learning

The most widely accepted view in the literature is that dopamine contributes to reinforcement learning. Learning to explore and exploit one's environment is critical for decision-making. To optimize decision-making, agents must identify and assign value to critical stimuli in a sensory-rich environment. Moreover, they must attribute associations (often causal) between relevant stimuli and key outcomes. This type of associative learning has been probed in animals using principal behavioral paradigms of conditioning.

Famous classical conditioning experiments conducted by Ivan Pavlov studying hungry dogs laid the foundation for various associative learning methods used in animals. Classical conditioning is a learning process which involves the pairing of a neutral conditioned stimulus (CS) like a bell, with an unconditioned stimulus (US) like food. Through associating this pairing, animals first show behavioral or physiologic responses to the food (US) but gradually develop these responses to the bell (CS) itself. Essentially, as the animal forms an association between the bell and food, it learns that the CS predicts the delivery of the US.

A plausible explanation for CS-US associative learning is the simple temporal relationship between the two. However reasonable this may be, a key experiment in the 1960s demonstrated that the temporal proximity of the CS and US pairing is not sufficient to drive learning. In this experiment, the learning of one CS blocked the learning of a novel CS. Animals were trained to associate one stimulus, a white noise (CS N) with a foot shock (US). After learning this contingency, CS N was presented with another stimulus, a light (CS L), to predict the same foot shock outcome. The compound presentation of both stimuli blocked the learning about the CS L association with the US. Essentially, since the CS N already predicted the US outcome, no new learning of the CS L was required be able to predict the US. Thus, this blocking effect provided the first evidence that learning required more than just temporally contiguous conditioned and unconditioned stimuli (Steinberg et al., 2013)(Kamin, 1969). Instead, the authors intuited that learning requires an element of surprise such that the US must be unpredicted in order for the animal to learn to ascribe meaning to the CS that preceded it.

Rescorla-Wagner Model

A few years after the discovery of the blocking effect, the Rescorla-Wagner model attempted to mathematically explain learning during classical conditioning. Two critical components to the model are 1) recognition that learning will only occur if there is a surprise, or a mismatch between the outcome and the agent's expected outcome and 2) on any given trial this expectation is based on the predictive value of all stimuli. The model asserts that the change in associative strength between a CS an US, or the amount of learning (ΔV), relates to the degree of unpredictability, or the difference between the actual US outcome (λ) and the expected outcome

(V). Conventionally, λ is set to a value of 1 to represent the presence of the US and 0 to represent the absence of the US. The term α is the rate at which learning occurs and relates to the saliency of the CS, but is assumed to be constant throughout learning (Rescorla & Wagner, 1972).

$$\Delta V = \alpha(\lambda - V)$$

The Rescorla-Wagner model demystified blocking by providing a mathematical rationale for why animals do not learn to associate a novel CS with an existing US. Using this model we see that presenting a novel CS (which holds no predictive value, $V=0$) together with a learned CS (which holds high/maximum predictive value, V=1) results in no change in associative strength (ΔV =0) because the sum of the two predictive values is 1, which essentially equals λ.

Ultimately, the model was very useful in understanding how discrepancies between expectations and outcomes, referred to as reward prediction errors, can lead to learning. Beyond the blocking phenomenon, the Rescorla-Wagner model is able to explain other facets of classical conditioning such as conditioned inhibition and overexpectation.

While extremely foundational, there are notable limitations with the model. For instance, it is not able to predict spontaneous recovery from extinction which can occur when an extinguished response returns after a rest period. Furthermore, the Rescorla-Wagner model cannot explain the latent inhibition phenomenon whereby preexposure to a CS that is first repeatedly presented without US and then is paired with US will result in slower acquisition than if it had been paired with the US from the very beginning. Finally, and perhaps most notably, Rescorla-Wagner is

limited to making predictions about learning based on discrete trials rather than capturing the multi-step, dynamic nature of learning (Niv & Schoenbaum, 2008).

Temporal Difference Reinforcement Learning Model

Whereas Rescorla-Wagner focuses on making predictions on a trial-by-trial level, the temporal difference reinforcement learning (TDRL) (Sutton, 1988) model aims to capture predictions on a multi-step time level. For example, in the context of classical conditioning, the Rescorla-Wagner model assumes that there are prediction errors only at the end of a trial once the actual outcome is weighed against an expected outcome. However, TDRL suggests that there are prediction errors at earlier timepoints and throughout the trial, like at the time the CS is delivered.

While there is considerable overlap between the Rescorla-Wagner and TDRL models, a crucial difference is that instead of predicting associative strength (ΔV), TDRL models the value of current and future reward on a continuous timescale. Value is defined as discounted, predicted future reward, where rewards in the proximal future are valued more than rewards in the distant future. Thus, value can be defined as the sum of all present and future rewards on a moment-to-moment basis, with a discounting factor (γ) for future rewards.

$$V(t)=r(t)+\gamma^1 r(t+1)+\gamma^2 r(t+2)\ldots$$

This can also be written as:

$$V(t)=r(t) + \gamma V\,(t+1)$$

Using this equation, we can estimate prediction error (δ) at time $t + 1$ as the discrepancy between the left and right sides of the equation. More intuitively, prediction error is the difference between the current prediction of future reward and the past prediction on the immediately preceding time step:

$$\delta = r(t) + \gamma V\,(t+1) - V(t)$$

The prediction errors that are generated at each moment are thought to drive learning as agents progressively minimize the discrepancy.

Dopaminergic contributions to learning: TDRL

Simulations of classical conditioning evaluating various types of stimulus representations have affirmed the TDRL model's predictions (Ludvig, Sutton, & Kehoe, 2012). However, in the 1990s, neurophysiological work conducted in behaving animals aimed to corroborate the mathematical theories experimentally. Work conducted by Wolfram Schultz recording dopamine neuron activity in monkeys found initial evidence of dopamine neuron bursts to cues predicting reward. Monkeys were placed in front of an occluded food box, and when the box opened, they would reach inside to receive a reward (Schultz, Apicella, & Ljungberg, 1993). The authors found clear evidence of

increased dopamine neuron activity to the door opening. However, it was not until a few years later that Read Montague noticed the predictive properties of these neurons and proposed that fluctuations in dopamine can be involved in predictive learning(Montague, Dayan, & Sejnowski, 1996). Then, seminal collaborative research conducted by Wolfram Schultz, Peter Dayan, and Read Montague uncovered a "neural substrate" of TDRL (Schultz, Dayan, & Montague, 1997). As monkeys trained on a classical conditioning task, the experimenters noticed greater dopamine firing to the CS than to the reward itself. Additionally, and perhaps more interestingly, unexpected reward deliveries (US in absence of CS) were associated with robust dopamine firing, while unexpected omissions of reward (CS in absence of US) were associated with pauses in dopamine firing. Essentially, this work found that dopamine neuron activity bidirectionally signaled reward prediction errors (RPEs) pertaining to future rewards.

Subsequent work has generally substantiated the role of dopamine in encoding prediction errors. The extent to which dopamine neurons encode error can be modulated by reward probability such that the magnitude of the reward response can vary monotonically, increasing as the probability of reward decreases (Fiorillo, Tobler, & Schultz, 2003). In addition to reward probability, reward magnitude has been known to regulate error signals with dopamine prediction errors varying according to the relative magnitude of reward, rather than absolute amounts (Tobler, Fiorillo, & Schultz, 2005). It has also been found that "belief" states can shape dopaminergic prediction errors when agents infer value with ambiguous information (Starkweather, Babayan, Uchida, & Gershman, 2017). While much of the reward prediction error literature focuses on positively valenced outcomes, unexpected aversive stimuli can also generate errors in burst firing of dopamine neurons, although some studies find positive responses (Kutlu, Tat, Christensen, Zachry,

& Calipari, 2023), some find negative responses (Ungless, Magill, & Bolam, 2004), while others find both (Matsumoto & Hikosaka, 2009).

Causal studies have furthered the evidence that dopamine is involved in reward error signaling. If dopamine provides a teaching signal, then it would follow that blocking dopamine neurotransmission should impair learning. Indeed, blocking dopamine neurotransmission at the D1 receptor with SCH 23390 before learning of an operant task, for example, impairs acquisition of the learned response, yet has no impact on performance at later stages of training once the response is learned (Choi, Balsam, & Horvitz, 2005). Interestingly, neither agonism nor antagonism of D2 receptors impacts performance or learning in a risky decision-making task (Stopper, Khayambashi, & Floresco, 2013). More recent causal studies leveraging optogenetic tools have allowed for temporally selective manipulations of dopamine dynamics to probe learning. For instance, phasic optogenetic stimulation of dopamine neurons is sufficient to induce conditioned place preference (Tsai et al., 2009) and intracranial self-stimulation(Witten et al., 2011). Even further, it has been shown that optogenetic stimulation of dopamine neurons in close temporal contiguity with a sensory cue can bestow meaning to the cue even in the absence of an actual reward. Subsequent presentations of the cue alone evoked endogenous dopamine neuron activity and conditioned behavior simultaneously (Saunders, Richard, Margolis, & Janak, 2018). Another notable experiment involves optogenetic activation of dopamine during the blocking phenomenon. As mentioned earlier, conditioning for a new stimulus (CS N) is blocked by an old stimulus (CS O) if the new stimulus does not confer any added predictive value. Because the CS O already reliably predicts the outcome, there is no new learning required for the CS N. The authors hypothesized that if they evoked an “artificial” error signal during reward delivery, it

would rescue the blocking effect of CS N. Consistent with this hypothesis, experimenters showed that stimulating dopamine neuron activity at reward delivery was sufficient to unblock the CS N when presented in compound with the CS O (Steinberg et al., 2013).

Dopaminergic contributions to learning: Alternate theories

Decades of research have supported the connection between dopamine and RPE. However, while RPE signaling has been the leading theory in the dopamine field, there are notable limitations to the TDRL model that have proven difficult to reconcile.

- Firstly, the findings that dopamine neurons respond to unexpected novel stimuli in the absence of value predictions (Horvitz, 2000)(Menegas, Babayan, Uchida, & Watabe-Uchida, 2017)(Kaźmierczak & Nicola, 2022) do not parsimoniously align with the TDRL model.
- Second, there is evidence that dopamine neurons can still signal error when expectations are violated but value is held constant (Chang, Gardner, Di Tillio, & Schoenbaum, 2017). For example, one study showed that individual dopamine neurons show stronger error responses to value-neutral changes in reward (switching from vanilla flavor to chocolate flavor of equal caloric and preferred value) than to value-based changes in reward (switching from small amount of vanilla to large amount of vanilla) (Takahashi et al., 2017).
- Third, while the literature is mixed, increases in dopamine neuron activity and release have been reported to aversive stimuli (Sorg & Kalivas, 1991)(Abercrombie, Keefe, DiFrischia, & Zigmond, 1989)(Brischoux, Chakraborty, Brierley, & Ungless, 2009) and classical fear

conditioning (Young, Joseph, & Gray, 1993)(Boulanger-Bertolus, Parrot, Doyère, & Mouly, 2021), aversive context conditioning (Fulford & Marsden, 2007), and negative reinforcement (Kutlu et al., 2021). According to the classic version of RPE theory, worse than expected outcomes should yield decreases in dopamine. However, in a recent study, mice trained to nose-poke to avoid shock showed dopamine responses to the foot shock that were positive and *increased* throughout training, instead of negative and decreasing as RPE would predict. Furthermore, dopamine responses to the cue predicting the shock remained consistent throughout learning, which also calls into question the predictions of TDRL (Kutlu et al., 2021).

- Fourth, fundamental to the TDRL model are predictions about future rewards, but this does not encompass learning about associations that lack an affective nature, as is the case in stimulus-stimulus learning, in which one stimulus predicts another stimulus which predicts reward (S → S → R). Say an auditory stimulus (AS1) predicts a visual stimulus (VS1) which predicts some reward (R1). Now, say another auditory stimulus (AS2) predicts another visual stimulus which predicts (VS2) which predicts another reward of equal value (R2). According to the TDRL, swapping visual stimuli such that AS1 → VS2 → R1, would yield no prediction error at the time VS2 is delivered because the predicted scalar value of the reward is the same in both VS1 and VS2 predictions, even though this violates expectations. Though the literature on dopamine in stimulus-stimulus learning is much more sparse, one group found that stimulating dopamine neurons during a preconditioning of a S1-S2 relationship could unblock learning about a later S2-reward relationship (Sharpe et al., 2017).

- Fifth, simple conceptions of TDRL predict tight timing between dopamine and behavioral dynamics during cue-reward learning: some believe that RPE updates the value signal to inform behavior in close temporal proximity to changes in dopamine. However, recent work using classical conditioning found that dopamine responses to the conditioned stimulus were observed days before the learned behavioral responses of anticipatory licking (Jeong et al., 2022). The interpretation here is either that dopamine is not signaling according to simple TDRL predictions or there are other processes that are slowing down the transition from cue-encoding to anticipatory licking. The latter does not entirely rule out TDRL.

Alternate models and frameworks to explain the manner in which dopamine influences learning have been proposed.

The Adjusted Net Contingency for Causal Relations (ANCCR), proposes that dopamine signals the inferred cause of meaningful outcomes. That is, rather than reflecting prospective predictions about future rewards as modeled by TDRL, dopamine signals retrospective causality as agents consider what cues preceded the reward (Jeong et al., 2022). However, due to the complexity of the model, it is not clear which aspects deviate from modern views of TDRL.

The Kutlu-Calipari-Schmajuk (KCS) model argues that an additional factor, perceived saliency, need be added to the RPE model. Perceived saliency takes into account the stimulus intensity and the attentional value of the stimulus (i.e., novelty). Adding this term to the RPE model accounts for positive dopamine responses in negative-valenced contexts while also ruling out alternate

explanations like ‘unsigned prediction error’ in which the directionality of the error signal is always positive regardless of the valence of the stimulus (Kutlu et al., 2021). This fortifies other theories proposing that psychiatric disorders, like substance use disorder, may be result of misalignment and misattribution of higher salience to drug-related stimuli (Kalhan, Redish, Hester, & Garrido, 2021).

A different model, the Adaptive Rate Cost of Performance to Reinforce (ACTR), proposes that learning can better be explained by direct policy learning than by value learning. ACTR suggests that rather than signaling error, dopamine contributes to modulating the rate of learning (Coddington, Lindo, & Dudman, 2023). There also exists a theory that dopamine broadly signals generalized prediction errors by combining sensory predictions and reward predictions (Gardner, Schoenbaum, & Gershman, 2018).

The micro-agent theory, informed by distributional reinforcement learning, suggests that rather than representing a single average probability of an expected outcome, dopamine represents a probability distribution. According to this idea, dopamine can reflect multiple, parallel future outcomes, and dopamine neurons can convey varying degrees of “optimism” about predictions. Therefore, rather than reversing from increased to decreased firing when outcomes are better than or worse than the distinct average expected outcome, the authors found that there is diversity in reversal points among dopamine neurons. This suggests that dopamine neurons can carry different value predictions (Dabney et al., 2020)(Kurth-Nelson & Redish, 2009).

Others have found evidence that dopamine can exert differential influence over learning and motivation depending on whether the dopamine signal is broadcast versus locally controlled, respectively(Mohebi et al., 2019). In this study, the authors compared VTA dopamine spiking activity with dopamine release in the NAc during decision-making. They found that while reward-associated cues led to increases in both spiking and release, only release correlated with dynamic reward expectations, which they presume is local, non-spiking mechanisms acting at dopamine terminals. Some work has even suggested that dopamine neuron heterogeneity may underlie some of the disparate findings such that some dopamine neurons encode values learning, while others encode motivation. There are accounts that NAc core may be involved in more motivational salience, while NAc shell may be more involved in motivational value processing(Bromberg-Martin, Matsumoto, & Hikosaka, 2010). The literature on regional differences is not unanimous, however. In fact, there is evidence of topographic subdivisions within the entire NAc that can significantly differ in their responses to cues predicting positive and negative valence stimuli (Lammel, Lim, & Malenka, 2014)(Liu et al., 2022), with some arguing that canonical RPE dopamine neurons project to more lateral subregions (de Jong et al., 2019).

While each of these frameworks has its own merits, their nascency still leaves many questions unanswered about how these theories may extend across various learning paradigms or decision-making contexts. For many of the recent models, there has not been sufficient time to vet each of the models by applying them to various behavioral datasets. Furthermore, the emphasis on using generally simpler behavioral tasks to test these theories means that the findings rarely embody the complexities of decision-making and are less likely to illuminate individual differences in value processing when there are limited resources and multiple choices. Ultimately, questions may still

remain about whether any existing single model will be able to explain all of the observed dynamics observed in the literature parsimoniously.

The gap

Although the theories disagree on the specific underlying computations, they generally agree that dopamine is a causal neurochemical contributing to learning. Presently, the TDRL model of dopamine has prevailed as the dominant theory to explain how animals learn to interface with predictive cues in their environment. While there is compelling computational and neurophysiological support for TDRL, one notable limitation is that it only accounts for predictions about *actual* rewards. As described earlier in great detail, learning and decision-making often involve reconciling the actual with the *counterfactual*. In the real world, there are often numerous competing options and possible outcomes, and in order for an agent to develop predictions about which cues or actions lead to specific outcomes, they often must also appreciate the alternate outcomes that could have been.

According to standard TDRL, prediction errors are the difference between current and past predictions of rewards at consecutive time steps. This allows agents to update predictions on the basis of outcomes that occurred in the past. This allows learning to occur through expectation violation, but this does not capture learning through outcomes that could have been worse or better than what actually happened. The extent to which an agent experiences regret or relief from an outcome may shape future predictions. It is also important to consider that there may be

heterogeneity in valuation systems across individuals that govern what types of strategies they employ when solving decisional problems.

These nuances reveal just how complex valuation can be and underscore the importance of developing intricate behavioral tasks with which to interrogate learning and decision-making systems. The experiments described in this thesis will investigate dopaminergic contributions to neuroeconomic decision-making both observationally and causally in mice. Leveraging an ethologically relevant foraging task, mice engage in valuation in the midst of competing offers for food rewards, allowing for more complex assessment of both factual and counterfactual features of the decision-making process. The following studies uncover distinct decision-making strategies that map onto dissociable dopamine dynamics. This work seeks to expand on the TDRL framework by asking how dopamine reflects rewards that were relinquished, how dopamine reflects value from the counterfactual past and the counterfactual future, and how dopamine relates to confidence in future predictions. The answers to these questions are critical for appreciating the parallel, and possibly separable, decision-making computations within the brain.

Chapter Two

Individual differences in decision strategy shape how mesolimbic dopamine encodes counterfactual future to influence choice confidence

2.1 Introduction

Counterfactual reasoning is the re-evaluation of paths untraveled during decision-making, as agents imagine alternate pasts or possible futures (Kishida et al., 2016; Van Hoeck, Watson, & Barbey, 2015), and relates to the confidence with which agents make decisions (Boldt, Schiffer, Waszak, & Yeung, 2019; Josephs, Larrick, Steele, & Nisbett, 1992; Roese & Olson, 1993; Sanna, Meier, & Turley-Ames, 1998; Zylberberg, Wolpert, & Shadlen, 2018). Dopamine in the nucleus accumbens (NAc) is a regulator of decision-making, implicated in motivation and reward prediction error (RPE) derived from evaluating actual outcomes (Aitken, Greenfield, & Wassum, 2016; Flagel et al., 2011; Hamid et al., 2015; Howe et al., 2013; Mohebi et al., 2019; Saunders et al., 2018; Schultz et al., 1997; Sharpe et al., 2017; Steinberg et al., 2013; Wassum, Ostlund, Loewinger, & Maidment, 2013b). But how does dopamine signaling represent *counterfactual* outcomes about rewards that could have been but were relinquished? And how does dopamine relate to the confidence with which decisions are made?

We monitored dopamine dynamics during internally-driven evaluation and re-evaluation processes, using an economic foraging task where mice have a limited time budget to spend seeking food rewards of varying subjective value. The self-paced nature of the task allowed for mice to evaluate and re-evaluate decisions by exhibiting change-of-mind behaviors, which have been linked to counterfactual processes (Coricelli et al., 2005; Frydman & Camerer, 2016; Steiner & Redish, 2014; Sweis, Thomas, et al., 2018). We found that dopamine tracked decision confidence during evaluation, and counterfactual values from the past and future. However, the nature of this encoding was different between individual animals as a function of the behavioral strategy used to increase their earnings. Optogenetic manipulation of mesolimbic dopamine release

altered evaluation and re-evaluation specifically in mice whose dopamine signal and behavior reflected future counterfactual values.

2.2 Materials and Methods

Animals

All procedures were approved by the Institutional Animal Care and Use Committee at the University of Minnesota. Experiments were performed using comparable numbers of both female and male mice(Shansky & Murphy, 2021). DAT-IRES-Cre transgenic mice(Bäckman et al., 2006) were originally obtained from The Jackson Laboratory (JAX Stock #006660), and maintained on a C57BL/6J genetic background by breeding in-house. Following stereotaxic surgery at 8-12 weeks of age, mice were singly housed on a 14:10 light:dark cycle.

Behavior

Pellet Training. Mice were introduced to flavored pellets 1 week prior to the start of their training. During this pre-training period, mice were transferred from regular rodent chow to a diet consisting of Bio-Serv full nutrition dustless precision pellets consisting of equal parts of chocolate, banana, grape, and plain flavored pellets (Bio-Serv product #F05301, #F0071, #F0079, #F07122). A free-feeding baseline weight was recorded as the average weight on three consecutive days of ad libitum pellet feed. Afterwards, daily feed was reduced by 0.5 g per day across four days. The day prior to beginning training, mice were introduced to the maze and given 15 minutes to explore the feeding sites. Each of the four feeders was filled with a specific flavor of pellets and surrounded by spatial cues to allow mice to become familiar with each restaurant.

Restaurant Row. Mice were run at the same time daily across consecutive days to maintain stable weights and motivational states. Training consisted of hour-long sessions of foraging on the maze in 4 stages. Stage 1 occurred on days 1-7 during which all offers were 1 second (associated tones = 4000 Hz, 500 msec). Stage 2 occurred on days 8-12 during which offers ranged from 1-5 seconds (associated tones ranged from 4000 – 5548 Hz, 500 msec). Stage 3 occurred on days 13-17 during

which offers ranged from 1-15 seconds (associated tones ranged from 4000 – 9418 Hz, 500 msec). Stage 4 occurred on days 18+ during which offers ranged from 1-30 seconds (associated tones ranged from 4000 – 15223 Hz, 500 msec). Offers were pseudorandomly selected such that all offer lengths were sampled and then reshuffled independently for each flavor. Each offer tone was presented when mice entered into the Offer Zone in a counter-clockwise direction, and repeated each second until mice either accepted or skipped the offer. After accepting the offers, the countdowns in the Wait Zone decreased in pitch 387 Hz steps per second in steps until reaching 0, at which point a uniquely-flavored pellet was dispensed using a Med Associates dispenser. Any pellets that were not consumed were flushed using mini-servos to prevent mice from returning to uneaten rewards at leisure. Audiotek speakers were placed by each restaurant to provide local sound. Behavioral tracking and programming were conducted using a Logitech HD Webcam positioned above the maze and AnyMaze software. Pre and post weights were taken for each animal and small portions of post-training feed were given to maintain body weights at ~80-85% of free feeding baseline. Photometry recordings and optogenetic manipulations were performed on the same maze as all training.

Fiber Photometry

Surgery. Under ketamine:xylazine anesthesia (cocktail 100:10 mg/kg), holes were drilled above the NAc core (AP +1.35, ML +/- 2.13; DV -4.3, at a 10 º from center angle) and VTA (AP -2.9, ML +/- 0.4; DV -4.55). Using a 33-gauge Hamilton syringe, 0.5 µL of AAV9-CAG-dLight1.3b (Addgene plasmid #125560; a gift from Lin Tian) and AAVdj-hSyn-FLEX-ChrimsonR (Addgene plasmid #62723; a gift from Edward Boyden) or AAV5-EF1a-DIO-eNpHR3.0 (UNC Vector Core) were injected at a rate of 0.1 µL/min bilaterally into the NAc core and VTA, respectively. After

allowing for sufficient viral diffusion for 10 minutes, the syringe was retracted slowly at a rate of 1mm/min. Fiber-optic cannula (400 μm, Doric Lenses: MFC_400/430-0.48_6mm_MF1.25_FLT) were implanted 0.05 mm dorsal to the injection site at a 10 degree angle targeting the NAc core and secured to the skull using jeweler's screws and cured dental resin (Geristore). Virus was incubated for at least four weeks to allow for sufficient expression of dLight and opsin transport to mesolimbic dopamine terminals.

Data Collection: Dopamine dynamics were measured using a Tucker Davis Technologies RZ5P fiber photometry processer. Blue (470 nm) and violet (405 nm) LEDs (ThorLabs) were modulated at distinct carrier frequencies (531 Hz and 211 Hz, respectively) for dLight excitation and isosbestic control. LED output power was maintained between 50-75 μW. Signals were filtered through a fluorescence mini cube (Doric Lenses) and measured with a femtowatt photodetector (Newport), sampled at 6.1 kHz. The distal end of the cable was coupled to a fiber optic patch cord (400 μm, 0.48 NA, Doric Lenses) which connected to fiberoptic ferrules implanted in animals. Recordings hemispheres were counterbalanced among animals (fig. S3). Behavioral sessions were aligned to fiber photometry recordings using TTL signals sent from Anymaze to the RZ5P.

Processing: All recorded signals were analyzed offline. Changes in dLight fluorescence were measured by fitting the 405 nm isosbestic signal to the 470 nm signal and calculating dF/F ([470 nm signal- fitted 405 nm signal]/ [fitted 405 nm signal]). The output of this processing step effectively corrected for photobleaching and was analyzed without any further rolling averages or smoothing. For behavioral event related analyses, signals were aligned to relevant behavioral timepoints and averaged within subjects across trials, then between subjects across days.

Optogenetic calibrations and validation

We used a five-port minicube (Doric Lenses) to filter excitation and emission channels for combined fiber photometry and optogenetic stimulation. For evoked recordings used for calibrations, a Master-8 (AMPI) was used to drive a 589 nm laser (Opto Engine LLC) to generate 0, 5, 10, 15, and 20 pulses at 20 Hz to stimulate VTA terminals in NAc expressing ChrimsonR, while simultaneously recording dLight1.3b responses through the RZ5P as described above. Five technical replicates were conducted for each parameter and an interval of 30 seconds was maintained between each stimulation. Laser power was set at 3-4 mW output to produce dopamine responses with amplitudes similar to the largest endogenous response observed after earning reward. Unique stimulation parameters were selected for each individual animal to ensure that dopamine responses were standardized across mice, and resembled physiologic levels of dopamine seen during task performance(Coddington et al., 2023; Markowitz et al., 2023; Pan, Coddington, & Dudman, 2021).

In order to validate terminal inhibition, a separate group of mice expressing eNpHR3.0 were delivered 2 seconds of continuous light (589 nm, 6-8 mW, Opto Engine LLC) at the time of earning a pellet on half of trials on two training days.

Optogenetic manipulations during behavior

Data Collection: After completing training, mice underwent a series of optogenetic manipulation testing days. An Anymaze Optogenetic Interface was coupled to a 589 nm laser (Opto Engine LLC) to interface between the behavioral software and light source. Closed-loop behavioral tracking through Anymaze allowed for laser activation immediately upon crossing into the Offer

Zone and termination at the time of accepting or skipping the trial. Mice received stimulation within the Offer Zone on 50% of offers above 15 seconds selected randomly across 1) all flavors, 2) less preferred flavors, and 3) more preferred flavors on different days. The order of these testing conditions was counterbalanced between animals. Stimulation days were interleaved with non-stimulation behavioral days. For photoinhibition of terminals during re-evaluation, the laser was activated upon mice crossing into the Wait Zone and terminated 4 seconds later or whenever the mouse quit, whichever occurred first.

Analysis

Probability of accepting, skipping, quitting, and earning were calculated as a proportion of all offers (trials), separated by flavor preference rank. Differences in skipping offers among behavioral phenotypes was quantified as the probability of skipping a 1s offer subtracted from the probability of skipping for a 30s and separated by rank. Thresholds were calculated each day, for each animal and each flavor. For Offer Zone thresholds, we fit a sigmoid to the binary Offer Zone outcome (accept or skip) as a function of offer and calculated the point of inflection (i.e., the offer at which a particular animal would shift from accepting offers to skipping offers for a specific flavor). Using a similar approach, we calculated Wait Zone thresholds, fit to the Wait Zone outcomes (earn or quit), and calculated the offer at which a particular animal would shift from earning to quitting for a specific flavor. OZDM and WZDM were categorized in groups according to either of two criteria: 1) a threshold decline probability >0.09 across training days 60-70 or 2) a change in temperature of threshold >0.01 (Fig 2).

Dopamine responses to offer were binned in short (1-5s), medium (6-15s), and long (16-30s) delay bins, and aligned to offer onset during mid-training (Days 30-37). Using a masking function, the data were restricted from the transition from the previous restaurant to the current trial's Offer Zone to constrain our analysis to offer-related dopamine transients. Responses post offer were then averaged (0 to 3s) and normalized to a pre-Offer period (-2 to -1s). Dopamine was then aligned to Offer Zone outcomes (accept and skip) and Wait Zone outcomes (earn and quit). Mean dopamine signals were calculated for accepts, skips, and quits by averaging dF/F during the pre-event period (-1 to 0s) or post-event period (0 to 2s) for earns, and normalizing to a preceding baseline (-2 to -1s).

Analysis of future value was conducted by calculating value remaining at the time of quit (Value Remaining (future value) = Threshold – Time Remaining in Countdown). Thresholds were calculated for each animal each day using the methods described above. Time remaining at countdown was defined as the number of seconds remaining in the countdown the animal would have had to wait to have earned reward and was calculated as Offer – Time to Quit. Peri-event dopamine signals were grouped in two bins: favorable future values (Value Remaining <0) and unfavorable future values (Value Remaining >0) and aligned to quits. Future time remaining was analyzed by splitting data into low (≤15s) and high (>15s) time remaining. Mean changes in dopamine to value remaining and time remaining were calculated by averaging the dopamine signal during the dips (-2 to 0s) and normalizing to a pre-dip baseline (-4 to -3s).

Figure 4: Absolute integrated angular velocity was calculated for each pass in the Offer Zone and used to signify IdPhi for each trial. Trials were split based on the z-score of their IdPhi values: zIdPhi>0 were classified as VTE and zIdPhi<0 were classified as NonVTE. IdPhi relates to path

curvature, so we utilized a path analysis previously reported in studies characterizing tortuosity of blood vessels in the retina (Hart, Goldbaum, Côté, Kube, & Nelson, 1999). Dopamine signals were aligned to the point of maximum curvature and separated based on trial outcome (accept or skip). Mean dopamine was calculated as the average dF/F peri-max curvature (-0.5 to 0.5s) and normalized to a preceding baseline (-4 to -2s).

zIdPhi was calculated for each pass in the Offer Zone and average zIdPhi on offer above 15 seconds were compared on stimulated and matched non-stimulated trials. Difference scores were calculated between stimulated and non-stimulated cases. Probability of quitting was calculated for each trial and probabilities were compared on trials that received Offer Zone stimulation and matched non-stimulated trials. Difference scores were calculated between stimulated and non-stimulated cases.

Immunohistochemistry

After behavioral testing, mice were deeply anesthetized using Beuthanasia (200 mg/kg, intraperitoneal) and transcardially perfused with ice-cold PBS and 4% paraformaldehyde in PBS for 15 minutes. Brains were extracted and post-fixed overnight in 4% paraformaldehyde in PBS at 4° C. Coronal sections were collected at 50 μm thickness using a vibrating microtome (Leica Microsystems) for staining with immunohistochemistry. Nonspecific binding was blocked with 2% normal horse serum + 0.05% Tween-10 + 0.2% Triton-X100 in PBS overnight at 4° C. After washing, slices were then incubated in mouse anti-GFP (Invitrogen, A-11120, diluted 1:500) rabbit anti-RFP (Rockland, 600-401-379, diluted 1:1000) primary antibodies diluted in the same blocking solution for 24 hours at 4° C. After four rinses in PBS containing 0.1% Tween-20,

sections were transferred to secondary antibody goat anti-mouse IgG Alexa 488 (Abcam, ab150115, diluted 1:1000) and donkey and rabbit Alexa 647 (Abcam, ab150075, diluted 1:1000) diluted in blocking solution for 24 hours at 4^{o} C. Sections were rinsed three times using PBS with 0.1% Tween-20 then mounted on slides and coverslipped using DAPI mounting (ProLong Gold antifade reagent with DAPI, Lot #250196) and imaged using fluorescence microscopy (Leica, BZ-X Series). Fiber optic tip locations were estimated using the Paxinos and Franklin's Mouse Brain Atlas (Paxinos & Franklin, 2001).

2.3 Results

Two types of decision-makers employ different economic strategies

To evaluate the role of mesolimbic dopamine transmission in economic decision-making, we trained mice of both sexes (Shansky & Murphy, 2021) in a foraging task (Sweis, Abram, et al., 2018) with a daily budget of one hour to obtain food rewards of four distinct flavors (Fig. 1a). As mice ran counterclockwise around the maze, they encountered a different Offer Zone at each corner. Upon entering the Offer Zone, mice were presented with a tone that signaled the delay the animal would have to wait before a reward was delivered. The delays were random, ranging from 1-30 seconds, and scaled with tone such that longer delays were signaled by higher-pitched tones. After evaluating tone presentation in the Offer Zone, mice could either skip the offer (continue to the next restaurant) or accept the offer (advance into the Wait Zone and begin the tone countdown). Mice earned a reward if they waited out the entirety of the delay, but at any point during the countdown, mice could re-evaluate their decision and quit the trial early by leaving the Wait Zone for the next restaurant. As mice were trained to evaluate longer offers, they developed economic strategies to increase earnings: increasing the number of laps run (Fig. 1b), decreasing the amount of time invested before quitting (Fig. 1c), and gradually switching from quitting to skipping behaviors (Fig. 1d). This ultimately resulted in increased earnings (Fig. 1e) and the development of flavor preferences over time (Fig. 1f), which scaled with the amount of time invested before quitting (Fig. 1g).

Interestingly, two behavioral phenotypes emerged with differing decision-making strategies (Fig. 2) Offer Zone Decision Makers (OZDM) showed a behavioral sensitivity to offer when evaluating decisions in the Offer Zone, accepting low-delay offers and rejecting high-delay offers (Fig. 1h).

In contrast, Wait Zone Decision Makers (WZDM) entered the Wait Zone more quickly (Fig. 3) and irrespective of offer, accepting low- and high-delay offers with similar probability (Fig. 1j). The increased probability of accepting low- versus high-delay offers scaled with flavor preference in OZDM (Fig. 1l), while no differences based on offer length or flavor were observed in WZDM (Fig. 1n). After accepting an offer and entering the Wait Zone, both OZDM (Fig. 1I) and WZDM (Fig. 1k) were more likely to re-evaluate their decision and quit while waiting out long delays versus short delays (Fig. 1m and 1o). These data suggest that while evaluating the decision to accept or skip an offer, OZDM were more sensitive to the future consequences of this decision. This was further evidenced by OZDM increasing their skipping behavior (Fig. 1p), as they learned to selectively accept offers that matched their willingness to wait (Fig. 1q), as demonstrated by their converging thresholds. In contrast, WZDM maintained constant levels of skipping throughout training (Fig. 1r). Offer Zone thresholds and Wait Zone thresholds never converged in WZDM, indicating that they continued to accept offers greater than their willingness to wait (Fig. 1s), and were thus less sensitive to the future consequences of their decisions.

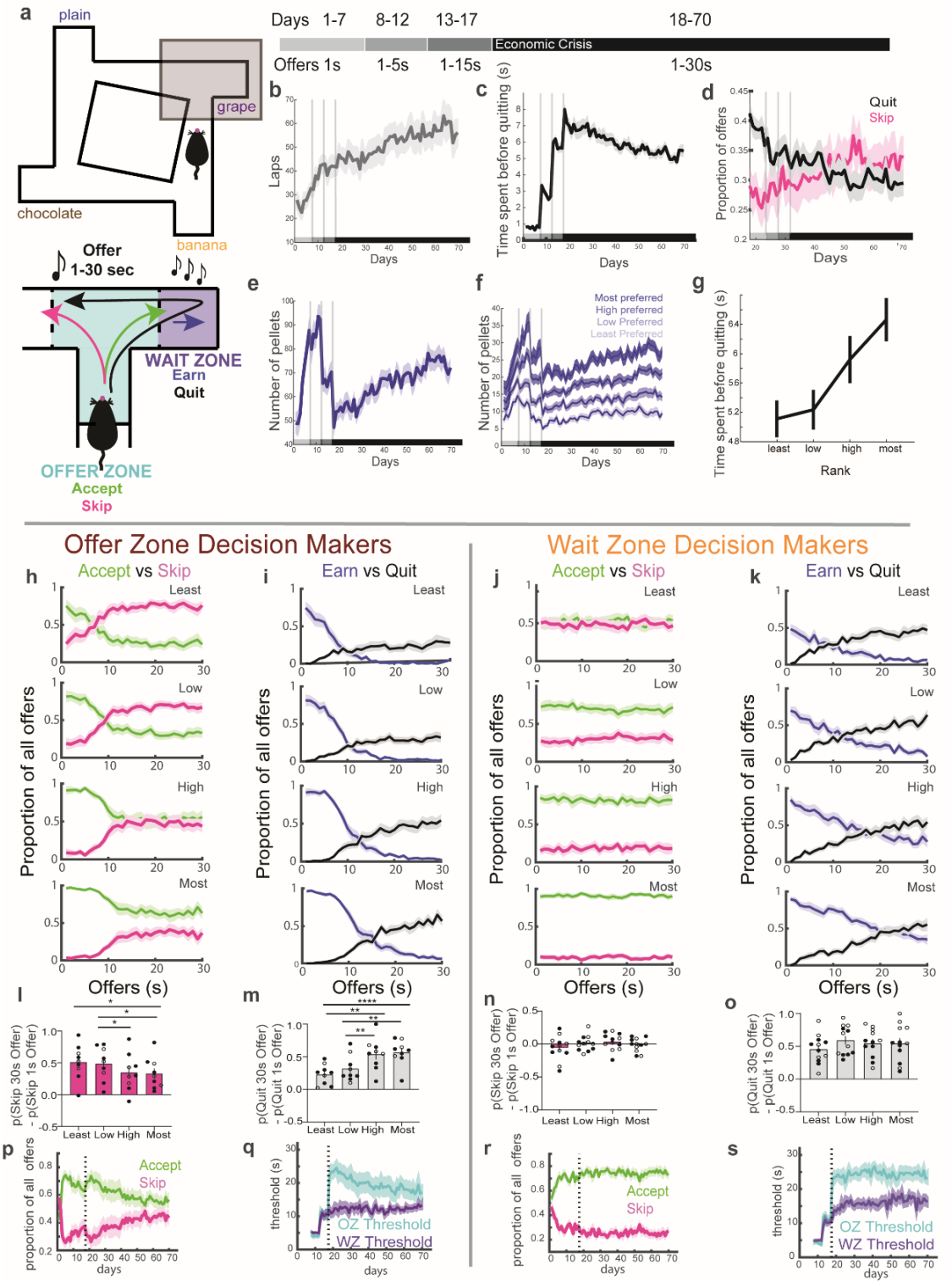

Figure 1 | Mice develop distinct economic decision-making strategies. (**a**) Task structure and training schedule. Top: maze with four flavors. Bottom: single restaurant showing Offer Zone, Wait Zone, and potential choices (**b**) Laps run. (**c**) Time spent before quitting. (**d**) Proportion of quits and skips. (**e**) Pellets earned. (**f**) Earnings by individual restaurant rank. (**g**) Time spent before quitting

by rank. (**h to k**) Proportion of offers accepted or skipped by rank and offer for OZDM (h) and WZDM (j), and proportion of offers quit or earned by rank for OZDM (i) and WZDM (j). (**l to o**) Difference in skip probability between high and low offers by rank for OZDM (l) and WZDM (n), and difference in quit probability between high and low offers by rank for OZDM (m) and WZDM (o). (**p to s**) Proportion of offers accepted or skipped across training for OZDM (p) and WZDM (r), and thresholds in Offer Zone and Wait Zone for OZDM (q) and WDM (s). Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *$p<0.05$, **$p<0.01$, ****$p<0.0001$ [ANOVA followed by Fisher's LSD post-hoc test]; see Data Table 1 for complete statistics.

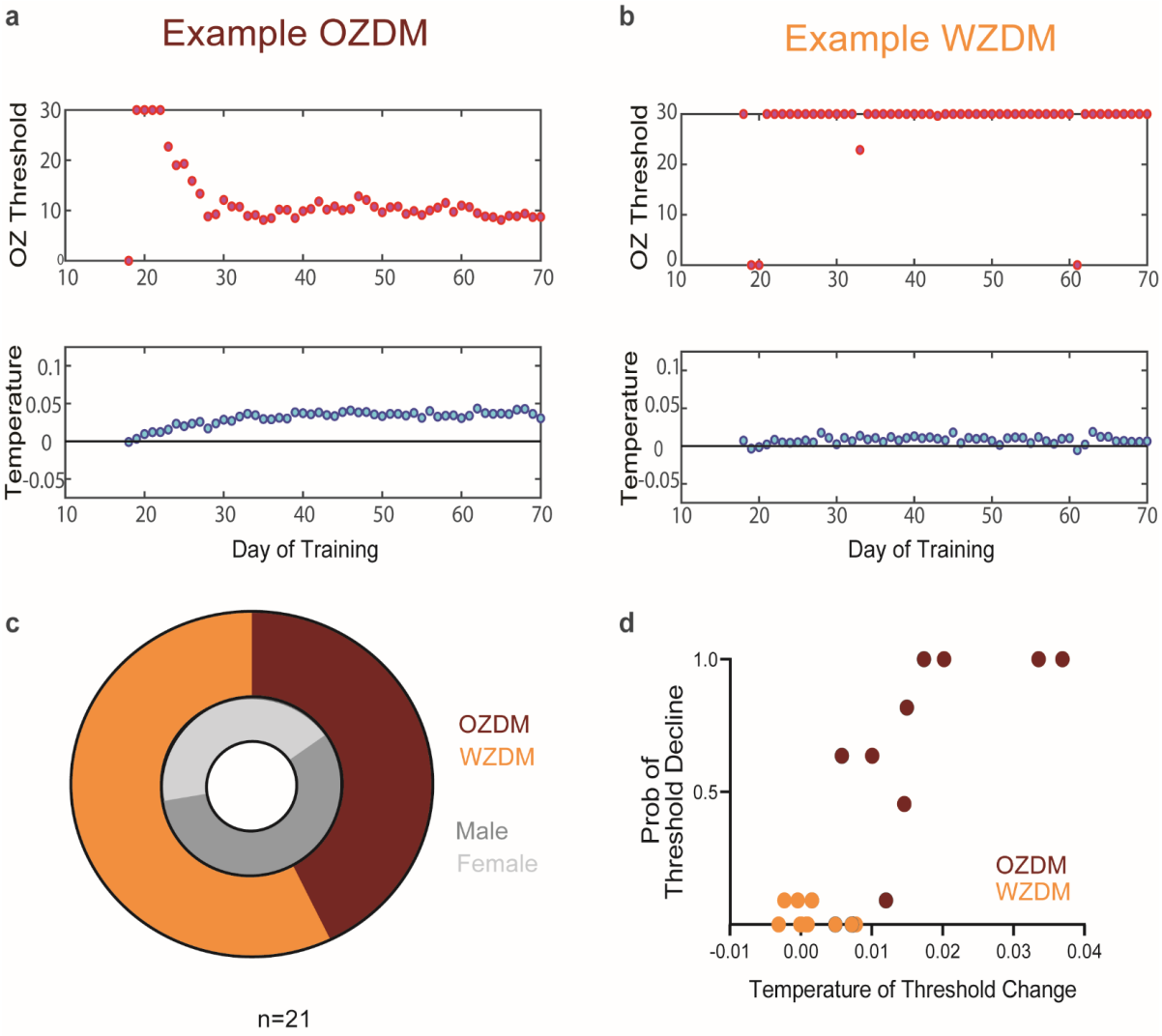

Figure 2 | Decision-making phenotype definition and characteristics. a) Top, Example decline in OZDM Offer Zone thresholds across training. Bottom, Example increase in temperature of Offer Zone Threshold change across training. **c)** Top, Example sustained Offer Zone thresholds in WZDM across training. Bottom, Example sustained temperature of Offer Zone Thresholds across training. **c)** Proportion of OZDM and WZDM by sex. **d)** OZDM and WZDM grouped according to either of two criteria: 1) a threshold decline probability >0.09 or 2) a temperature of threshold change >0.01.

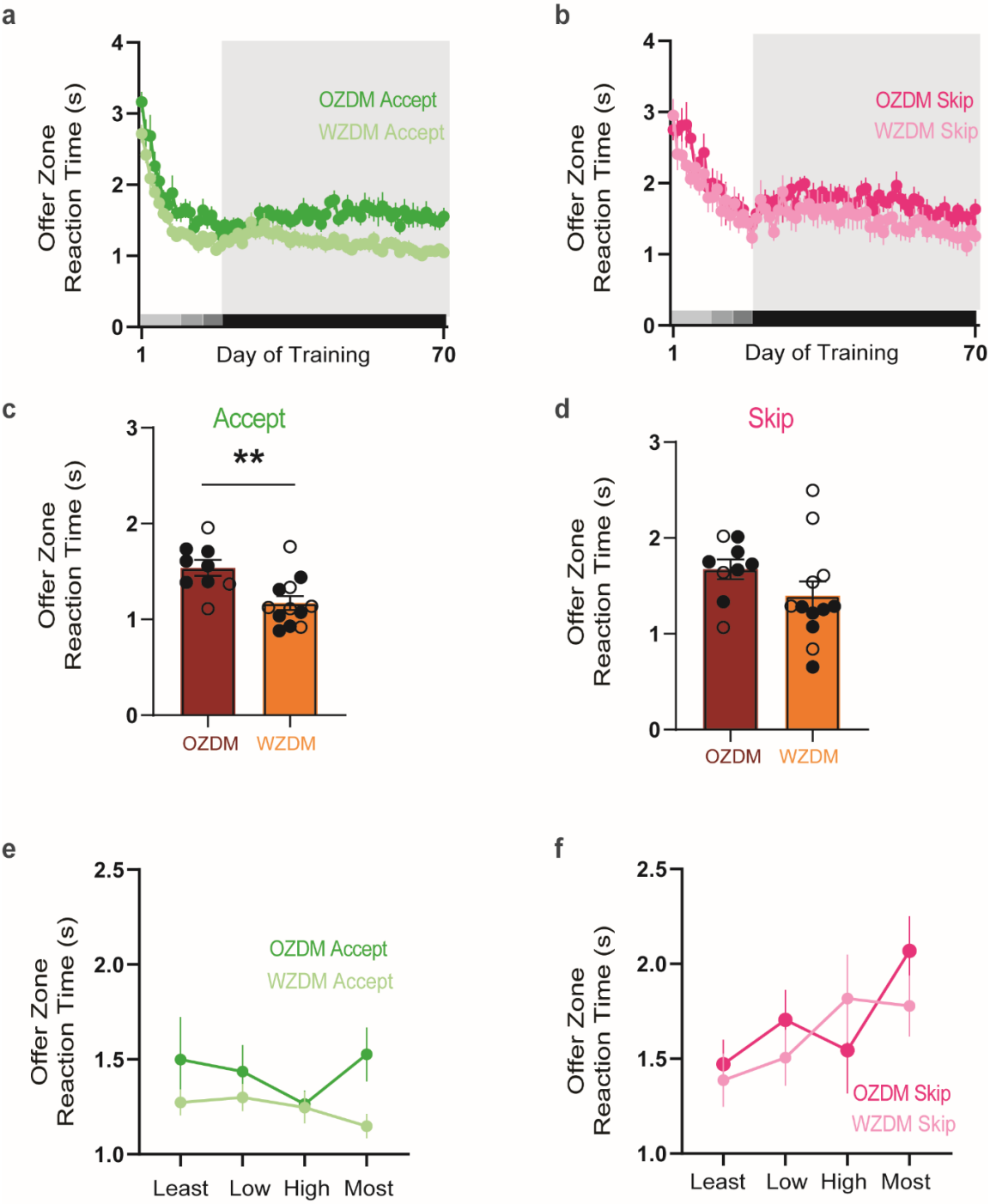

Figure 3 | OZDM and WZDM reaction times during offer evaluation. a) Reaction times to accept an offer in Offer Zone across training. **b)** Reaction times to skip an offer in Offer Zone across training. **c)** Average reaction times in the offer zone on accepted trials and **d)** skipped trials. **e)** Reaction times to Accept offer by flavor rank **f)** Reaction times to Skip offer by flavor rank. Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. **p<0.01 [ANOVA main effect]; see Data Table 1 for complete statistics.

Dopamine dynamics reflect decision strategy

To determine how these decision-making strategies were related to mesolimbic dopamine dynamics, we used fiber photometry to monitor dopamine signals in the NAc. We infused AAV9-CAG-dLight1.3b into the NAc core to express the dLight1.3b fluorescent biosensor (Mohebi et al., 2019; Patriarchi et al., 2018) (Fig. 4a), and implanted bilateral fiber optics within the NAc core (Fig. 4b-c and Fig. 5). These experiments were conducted using DAT-Cre transgenic mice, and we infused a second virus (AAVdj-hSyn-DIO-ChrimsonR) into the VTA to express a red-shifted excitatory opsin(Klapoetke et al., 2014) in a Cre-dependent fashion (Fig. 4d-e). This experimental design allowed us to stimulate mesolimbic axon terminals with red light, through the same optic fiber used to monitor dLight fluorescence with blue light(Lefevre et al., 2020) (Fig. 4f).

Early in training, OZDM and WZDM showed comparable dopamine responses after earning a pellet (Fig. 6). After offer delays reached 1-30 seconds, we compared dopamine signals aligned to the onset of offers with a short delay (1-5 s), medium delay (6-15 s), or long delay (15-30 s). The dopamine response was inversely related to offer length in both OZDM (Fig. 4g-h) and WZDM (Fig. 4i-j), indicating dopamine signals scaled with delay to reward (offer), and implying a neurochemical representation of expected cost in both groups. The average dopamine signal showed a transient peak once per second in both groups, corresponding to each individual tone presentation during the offer, and providing further evidence that OZDM and WZDM could both perceive the offer tone. Importantly, these data indicate that neural signals in WZDM reflected offer-based valuation, even if their behavior did not.

We then analyzed dopamine dynamics within the Offer Zone by separating decisions to accept or skip. OZDM showed bidirectional dopamine responses while evaluating decisions and their future

consequences, with dopamine increasing prior to accepting and decreasing prior to skipping an offer (Fig. 4k-m). WZDM also showed an increased dopamine signal prior to accepting an offer, but no decrease in signal on skip trials (Fig. 4m-o). We uncovered bidirectional dopamine dynamics in the Wait Zone in both groups based on outcome (earn versus quit). When both groups waited through the entire countdown and earned a pellet, we observed robust increases in dopamine at the time of pellet delivery (Fig. 4p, s). Dopamine also tracked the countdown itself, with small increases in dopamine occurring each second as individual countdown tones were presented. Conversely, when OZDM and WZDM re-evaluated their decision and quit the countdown early, there was a decrease in dopamine immediately preceding the quit (Fig. 4q, t). Both the increase in dopamine following earns and decrease in dopamine preceding quits were exaggerated in WZDM relative to OZDM (Fig. 4r). Dopamine signals were thus more robust and dynamic in the Offer Zone for OZDM and in the Wait Zone for WZDM, highlighting a correspondence between decision-making strategy and dopamine signaling in the zone where each mouse made its decision.

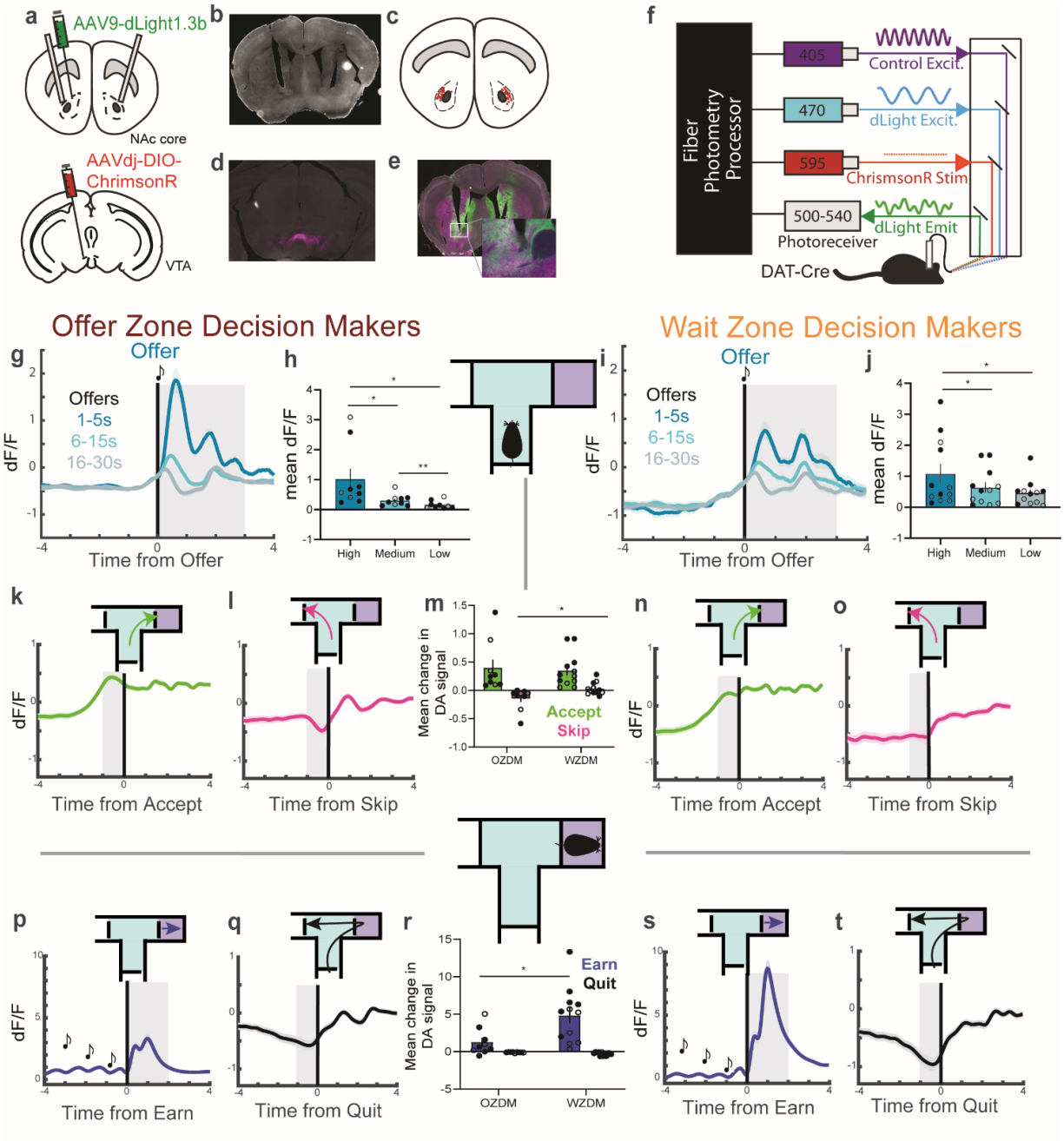

Figure 4 | Mesolimbic dopamine dynamics reflect decision location. (**a**) Schematic of virus transfection sites in NAc (top) and VTA (bottom). (**b**) Example of fiber optic placements. (**c**) Fiber optic tip placements for all animals. (**d**) Example histology showing viral expression of ChrimsonR in VTA (magenta). (**e**) ChrimsonR terminal expression (magenta) and dLight1.3b expression (green) in NAc. (**f**) Fiber photometry recording setup. (**g to j**) Time course of dopamine response to offer in OZDM (g) and WZDM (j), along with average response to high, medium, and low delay offers in OZDM (h) and WZDM (j). (**k to o**) Time course of dopamine responses to accepted offers in OZDM (k) and WZDM (n); skipped offers in OZDM (l) and

WZDM (o); and average responses (m). (**p to t**) Time course of dopamine response to earning a pellet in OZDM (p) and WZDM (s); quitting in OZDM (q) and WZDM (t); and average responses (r). Shaded gray boxes indicate time windows used for quantification. Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *p<0.05, **p<0.01 [ANOVA followed by Fisher's LSD post-hoc test]; see Data Table 1 for complete statistics.

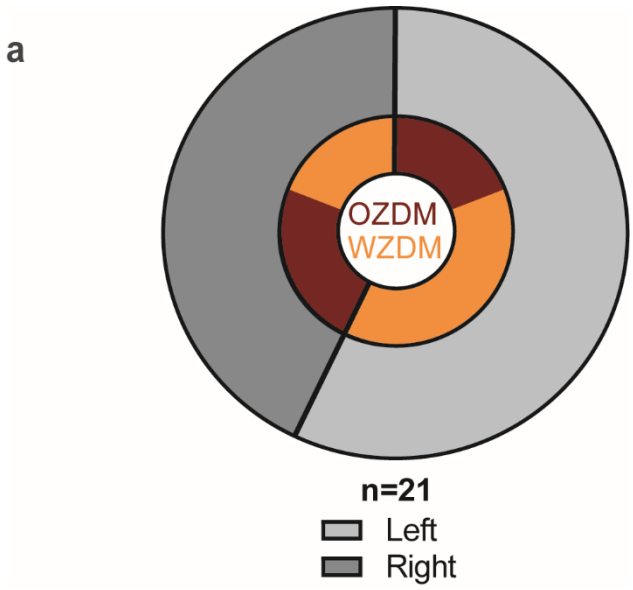

Figure 5 | Recording location for fiber photometry. Recordings from both left and right hemispheres were conducted in OZDM and WZDM

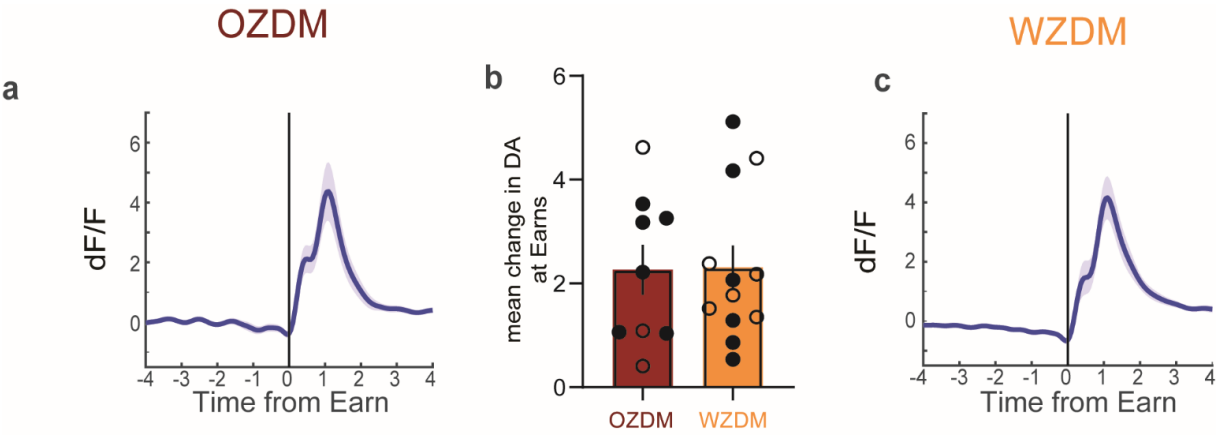

Figure 6 | Dopamine signals early in training before rewards become economically expensive. Responses were comparable in OZDM (a) and WZDM (b).

A specific association between dopamine dips and future counterfactual value in OZDM

After mice choose to accept an offer and enter the Wait Zone, they have an opportunity to re-evaluate their decision while waiting during the countdown to pellet delivery. Quitting behaviors arise from re-evaluation of choices, and thus represent a form of counterfactual reasoning about what alternate outcomes could have been (Sweis, Thomas, et al., 2018). If mice are re-evaluating an offer and concluding that it was worse than what could have been had they made a different decision, we would expect a decrease in dopamine signal from this internal re-evaluation process, a pattern that was evident in both OZDM and WZDM immediately preceding quits (Fig. 4, q and t). Importantly, there were no external stimuli that informed the mice to quit or altered expectations– this behavior was entirely internally motivated. Thus, decreases in dopamine during quitting on this task align more closely with a counterfactual process of what could have been had the animal made a different decision, and cannot be attributed to simple disappointment arising from expectation violation, out of the animal's control.

To further determine what information dopamine signaled before and after quits, we examined dopamine dynamics in relationship to the economic past and future. In this context, the future can be characterized as the future time remaining in the countdown after quitting, or more abstractly as the future value remaining in the countdown after quitting, which we can calculate by subtracting the time remaining in the countdown at the time of quit from the animal's willingness to wait (threshold). Economically-favorable quits occur when the time remaining in the countdown exceeds the animal's threshold (value remaining < 0), because the animal is relinquishing a future that requires waiting longer than it is typically willing to wait. Thus, favorable quits exemplify

consideration of the future value of an imagined prospective reward. Conversely, economically-unfavorable quits occur when the time remaining in the countdown is less than the animal's threshold for that reward (value remaining > 0), because the animal would normally have been willing to wait out the remaining time (Fig. 7a). Interestingly, we found that OZDM and WZDM displayed different proportions of favorable and unfavorable quits. OZDM had higher rates of favorable quits that increased across days during training (Fig. 7b), while WZDM had higher rates of unfavorable quits which remained relatively constant throughout training (Fig. 7c). This further supports the notion that OZDM were more sensitive to the future consequences of their decisions than WZDM.

Dopamine dynamics before quitting also differed between OZDM and WZDM mice. WZDM showed dips in dopamine for both favorable and unfavorable quits, while OZDM showed dips in dopamine specifically in cases of favorable quits, where the value of the future imagined reward was low (Fig. 7d-g). Importantly, these quit-related dopamine dips were not related to future time remaining in the countdown in either OZDM or WZDM (Fig. 7h-k). Furthermore, when we controlled for time remaining until earning a pellet, we found that dopamine was significantly higher during the countdown in high-value offers than low-value offers (Fig. 8a-c), providing further evidence that dopamine tracked value more closely than time.

To assess the specificity of the relationship between dopamine dynamics and counterfactual future, we also considered counterfactual past. If quitting is a re-evaluation resulting in a change-of-mind decision, then recovery from this re-evaluation after quitting may produce a concomitant rebound in dopamine. To test this, we examined dopamine dynamics after quitting as a function of past time (time invested into the countdown prior to quitting) and past value (difference between the

offer received and the animal's willingness to wait; Fig. 9). High-value trials are those where the animal has a high propensity to wait (i.e., more preferred flavors) and offer delay is short (i.e., low cost), whereas low willingness to wait and long offer delays constitute low-value trials. After quitting, rebounds in dopamine scaled inversely with past value in both OZDM and WZDM: low-value trials elicited the largest rebounds in dopamine, whereas high-value trials elicited the smallest rebounds in dopamine (Fig. 9b-e). This suggests that dopamine rebounds after quitting represent consideration of past value. This effect was not seen when separating trials by past time (Fig. 9f-i), implying these rebounds in dopamine after quits selectively represented past value.

Dopamine dynamics after the quit were thus reflective of a counterfactual *past* value taken in both groups, while differences in dopamine dynamics before the quit were specific to counterfactual *future* value loss in OZDM only. These results provide further evidence that OZDM considered the future in a way that was distinct from WZDM.

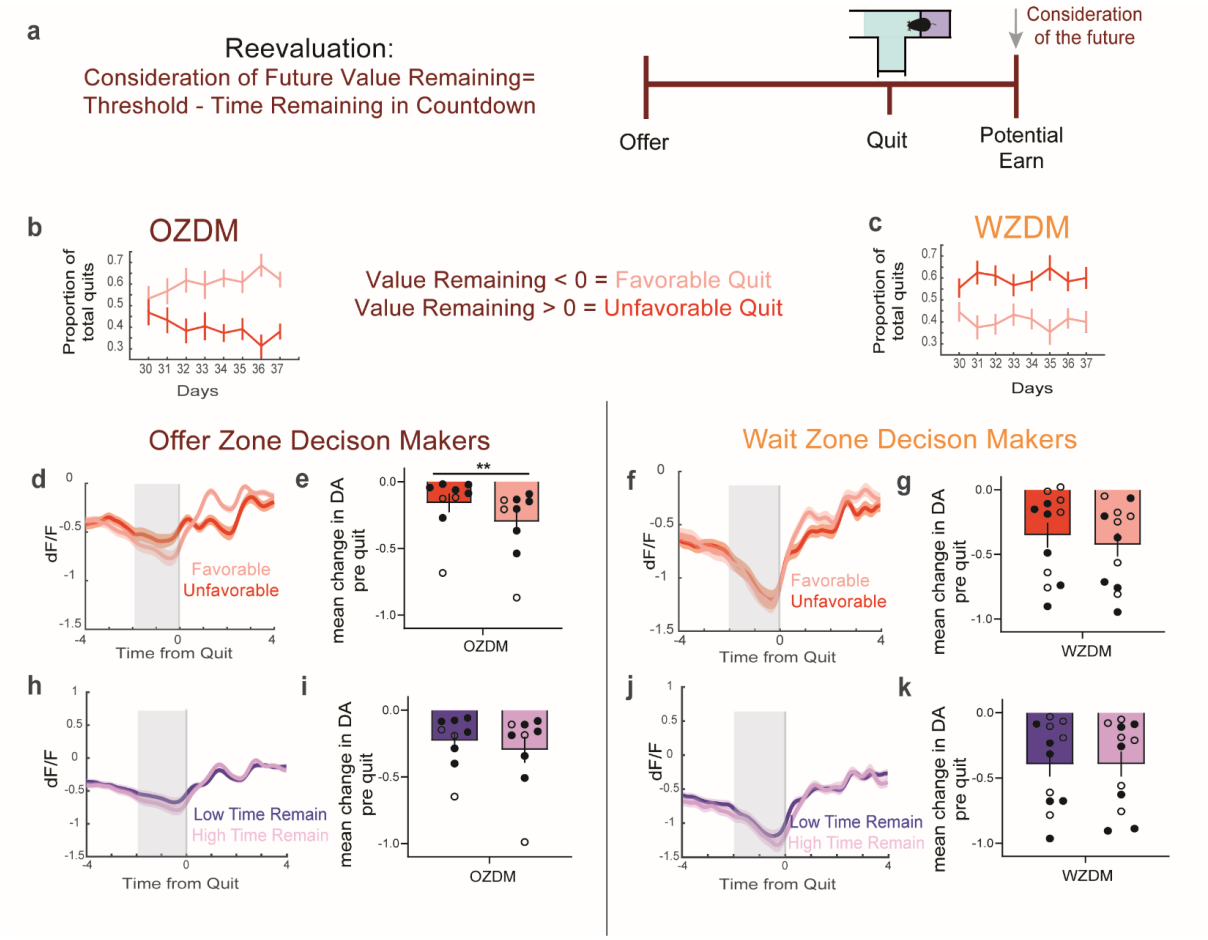

Figure 7 | Dopamine dips prior to quitting represent future value remaining in OZDM. (**a**) Schematic conceptualizing future value: the difference between willingness to wait (threshold) and time remaining in the countdown at quit. Values greater than zero mean it would be economically unfavorable to quit, while values less than zero mean it would be economically favorable to quit. (**b to c**) Favorable and unfavorable quits in OZDM (b) and WZDM (c). (**d to g**) Dopamine dynamics during favorable and unfavorable quits in OZDM (d) and WZDM (f), along with mean change in OZDM (e) and WZDM (g). (**h to k**) Dopamine dynamics while quitting with low and high time remaining in countdown in OZDM (h) and WZDM (j), along with mean change in OZDM (i) and WZDM (k). Shaded gray boxes indicate time windows used for quantification. Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *p<0.05, **p<0.01 [ANOVA followed by Fisher's LSD post-hoc test]; see Data Table 1 for complete statistics.

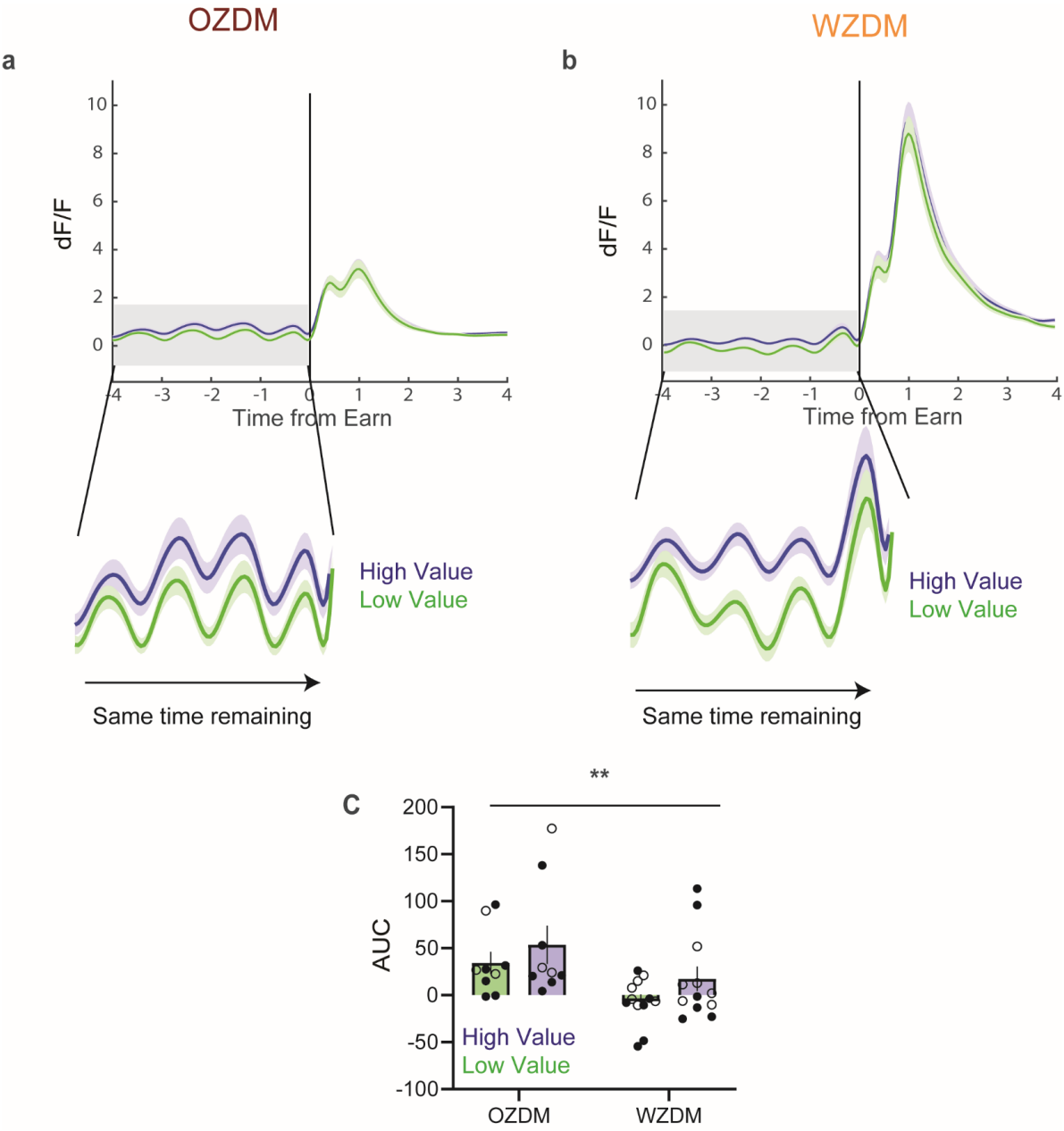

Figure 8 | Dopamine reflects value and not time to earn a) OZDM and **b)** WZDM display elevated dopamine during the countdown on higher value offers despite identical time to reward. **c)** Area under the curve for high- and low-value offers over equal windows of time to reward (4 seconds). Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. **p<0.01 [ANOVA main effect of value]; see Data Table 1 for complete statistics.

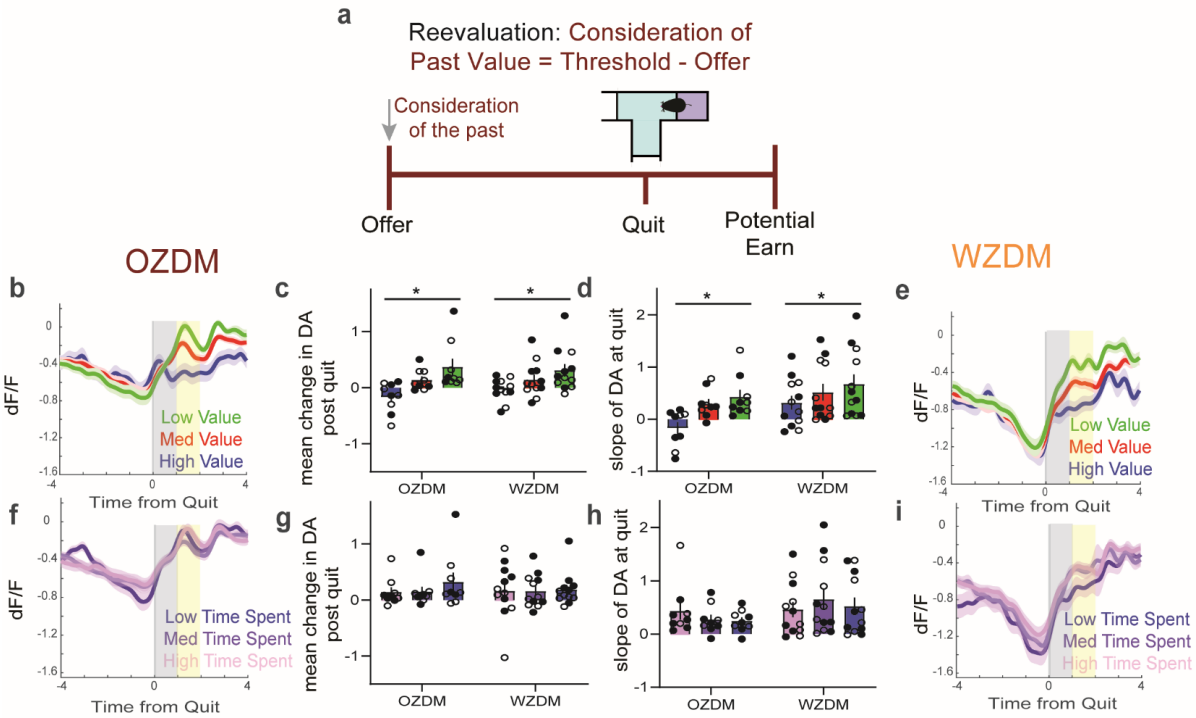

Figure 9 | Dopamine rebounds after quitting represent past value in both OZDM and WZDM. **a**) Schematic conceptualizing past value: the difference between willingness to wait (threshold) and offer delay. Highest values refer to short delays at more preferred flavors, while lowest values refer to long delays at less preferred flavors **b to e**) Dopamine dynamics during quit rebounds in OZDM (b) and WZDM (e), along with mean change in yellow shaded window (c) and slope in gray shaded window (d). **f to i**) Dopamine dynamics after quitting with low and high time spent in countdown in OZDM (f) and WZDM (i), along with mean change in yellow shaded window (g) and slope in gray shaded window (h). Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *p<0.05, **p<0.01 [ANOVA followed by Fisher's LSD post-hoc test]; see Data Table 1 for complete statistics.

Dopamine dynamics in OZDM relate to decision confidence during evaluation

In this task, consideration of the future also occurs during evaluation of decisions in the Offer Zone, where mice exhibit reorientation behaviors referred to as vicarious trial and error (VTE). VTE is a well-established behavior in rodents and correlates with future planning and deliberation (Papale, Zielinski, Frank, Jadhav, & Redish, 2016). During VTE in rodents, neural representations of possible future outcomes sweep serially along different potential paths and alternate between goals, suggesting that animals are considering potential options (Johnson & Redish, 2007; Kay et al., 2020; Papale, Stott, Powell, Regier, & Redish, 2012; Papale et al., 2016; Redish, 2016; Steiner & Redish, 2012; Stott & David Redish, 2014). VTE is best captured by calculating the integrated angular velocity (IdPhi) within the Offer Zone, which relates to the curvature of the path the animal takes to either accept or skip an offer. We defined VTE events as those with high IdPhi values (zIdPhi>0), which exemplified more variable paths with greater tortuosity (Fig. 10a), and correlated with higher degrees of deliberation on this version of the Restaurant Row task (Sweis, Thomas, et al., 2018).

As training progressed, OZDM began to demonstrate more VTE during evaluation than WZDM (Fig. 10b). This was indicated by a rightward shift in the distribution of IdPhi (Fig. 10c), a change in the distribution of VTE and non-VTE trials (Fig. 10d), greater maximum path curvature (Fig. 10e), and increased overall path curvature (Fig. 11a-b). Together, these results suggest that OZDM engaged in more deliberation during evaluation, as they considered future consequences of their decisions. As expected, path curvature was greater on VTE trials than non-VTE trials in both groups (Fig. 10f). Mice were more likely to make favorable quits after accepting offers on VTE trials than on non-VTE trials (Fig. 10g). This suggests that decisions were made

with a lower degree of confidence on VTE trials, allowing for quicker re-evaluation and change-of-mind correction, while value remaining was still low.

We next determined the extent to which VTE influenced dopamine dynamics during evaluation in the Offer Zone. We examine dopamine dynamics aligned to the time of peak path curvature to capture deliberative events (Fig. 10a, red squares) as mice made decisions to accept or skip offers, separating trials based on the presence or absence of VTE (Fig. 10h-k). Interestingly, when OZDM exhibited VTE before accepting an offer, we observed a smaller peak signal compared to accepted offers without VTE (Fig. 10l). Thus, as OZDM evaluate the future consequences of their decisions, dopamine levels positively correlated with decision confidence in OZDM. In contrast, WZDM showed comparable dopamine levels on trials with or without VTE (Fig. 10n). In both OZDM and WZDM, dopamine dynamics were similar following decisions to skip on trials with or without VTE (Fig. 10, m and o). While these analyses were conducting on data collected after extended training (Fig. 10b, shaded area), similar patterns in dopamine dynamics were also apparent earlier in training (Fig. 12a-e).

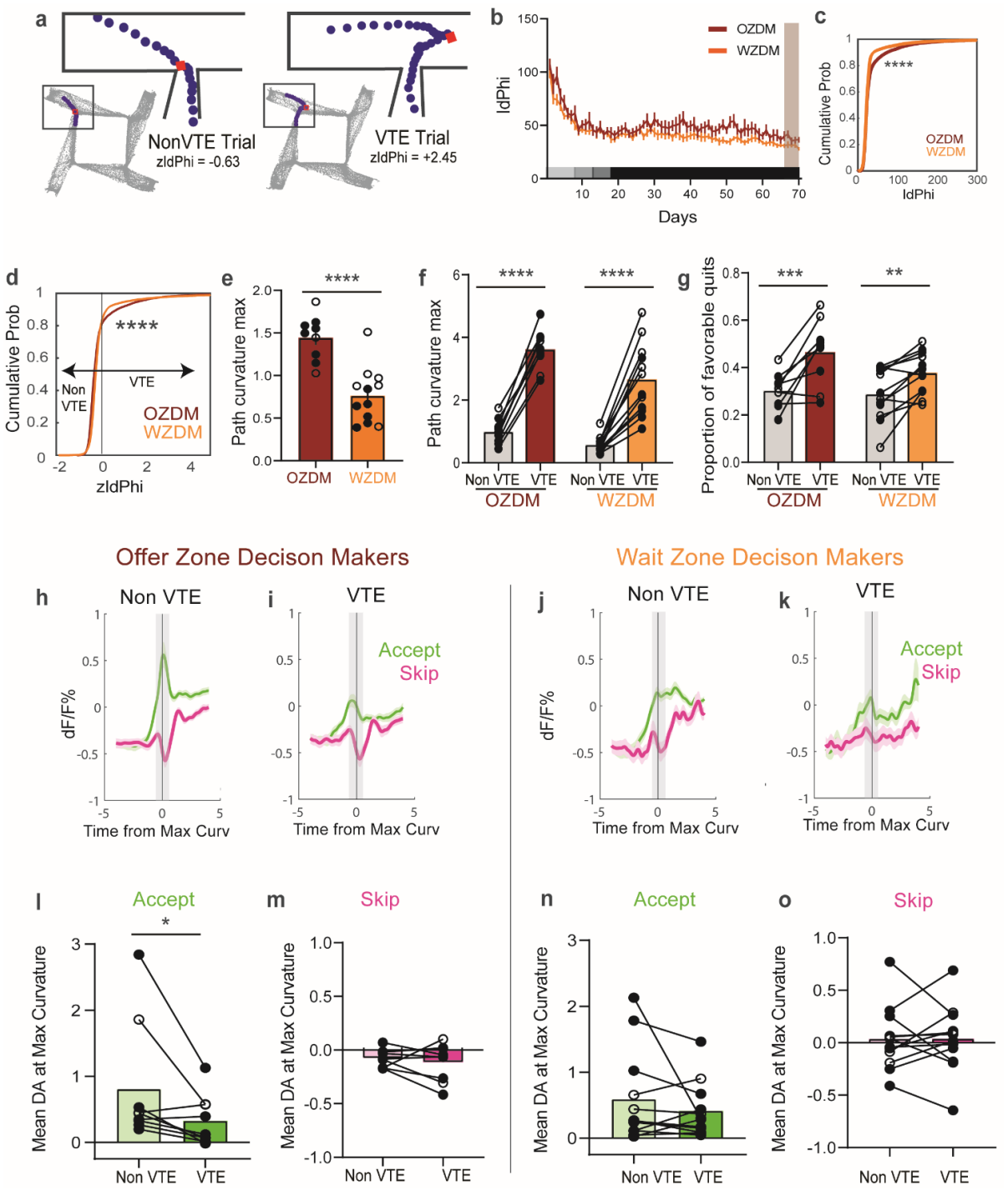

Figure 10 | Dopamine dynamics in OZDM relate to decision confidence during evaluation. (**a**) Example path curvatures on trials with low zIdPhi (Non-VTE, left) and high zIdPhi (VTE, right); red square represents point of maximum curvature. (**b**) IdPhi across training; dark gray shading represents dopamine recording days. (**c to d**) Distribution of IdPhi (c) and zIdPhi (d) (**e**) Maximum path curvature. (**f to g**) Maximum path curvature (f) and proportion on favorable quits (g) on Non-VTE and VTE trials. (**h to j**) Dopamine dynamics aligned to point of maximum path curvature in OZDM on Non-VTE (h) and VTE (i) trials, and WZDM on Non-VTE (j) and VTE (k) trials. (**l to o**) Mean dopamine response in gray

shaded window for Non-VTE and VTE trials that were accepted (l) or skipped (m) in OZDM, and accepted (n) or skipped (o) in WZDM. Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *p<0.05, **p<0.01 [Kolmogorov-Smirnov test (c to d) or ANOVA followed by Fisher's LSD post-hoc test]; see Data Table 1 for complete statistics.

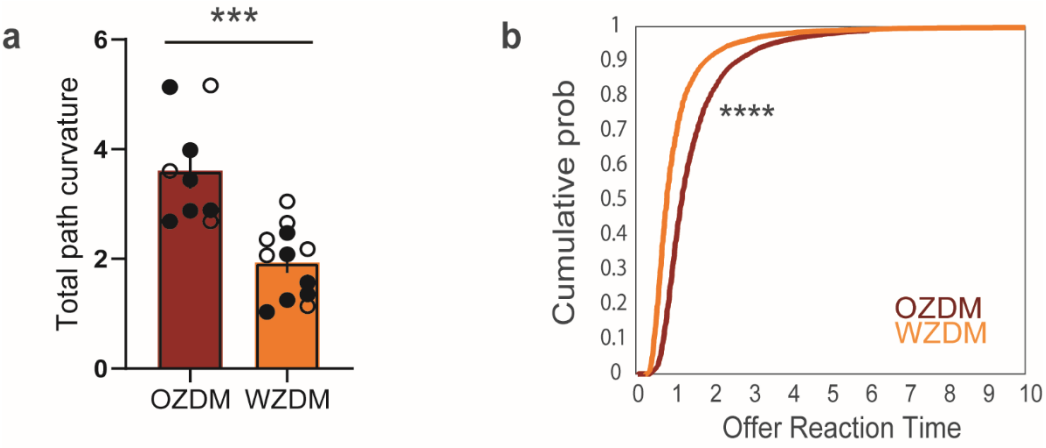

Figure 11 | Total path curvature and reaction times during evaluation a) WZDM have lower overall sum of path curvature and **b**) faster reaction times than OZDM. Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. ***p<0.001, ****p<0.0001 [ANOVA main effect (a) or Kolmogorov-Smirnov test (b)]; see Data Table 1 for complete statistics.

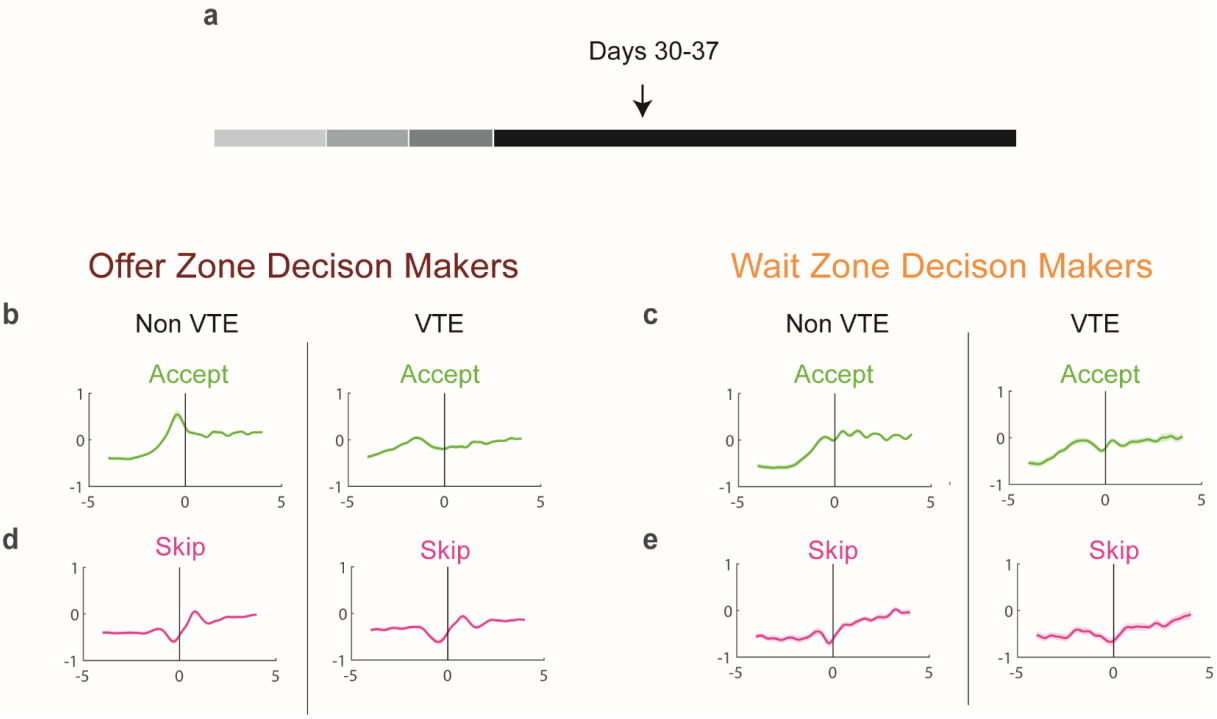

Figure 12 | Dopamine dynamics relate to decision confidence earlier in training follow similar trends as late training. a) Mid training recording days. **b)** OZDM and **c)** WZDM dopamine dynamics during accepts aligned to time at which maximum path curvature occurs on Non-VTE (left) and VTE (right) trials; **d)** OZDM and **e)** WZDM dopamine dynamics during skips aligned to time at which maximum path curvature occurs on Non-VTE (left) and VTE (right) trials.

Optogenetic enhancement of dopamine influences evaluation and re-evaluation of decisions in OZDM

Since our previous analyses suggested that OZDM displayed more deliberative and future-valuing behaviors than WZDM when evaluating and re-evaluating decisions, we wanted to determine if manipulating dopamine dynamics could differentially impact evaluation and re-evaluation of decisions in a strategy-specific manner. Using the same cohort of mice after completion of training, we delivered bilateral light stimulation (589 nm, 20 Hz) to mesolimbic dopamine terminals expressing ChrimsonR (Fig. 13a). Since ChrimsonR expression varied between animals, we leveraged expression of dLight in the NAc to construct stimulation-response curves for each individual animal. We then tailored optogenetic stimulation parameters for each animal to standardize the evoked signal across mice and to produce responses in a range ~3-10% dF/F, similar to endogenous signals (Coddington et al., 2023; Markowitz et al., 2023; Pan et al., 2021) for offer presentations and pellet consumption (Fig. 13b-c and Fig. 14).

We delivered optogenetic stimulation of dopamine terminals in the Offer Zone on half of all offers >15 sec (Fig. 13d). Stimulation was initially delivered for only the two less preferred flavors, where probability of offer acceptance is normally low. This manipulation reduced deliberative behaviors (zIdPhi and Offer Zone reaction time; Fig. 15) in OZDM but not WZDM, compared to control trials of the same type during the same session with no stimulation (Fig. 13, e and h). On a separate day of testing, we repeated this experiment but delivered stimulation on half of all offers >15 sec for the two more preferred flavors, where probability of offer acceptance is normally high. This manipulation had no significant effects in either group (Fig. 13, f and i), ruling out the possibility that stimulation of dopamine release simply alters the motor path of the animal. This interpretation was further reinforced when we delivered stimulation for all flavors, which did not change deliberative behaviors in either group (Fig. 13, g and j). The lack of effect

in WZDM with identical stimulation conditions suggests that stimulation of dopamine release alters decision-making in a strategy-specific manner, possibly by changing the deliberative process by which OZDM evaluate the future consequences of accepting an offer.

After accepting offers, mice re-evaluated their decisions in the Wait Zone. With prior stimulation for the two less preferred flavors, OZDM decreased their probability of quitting (Fig. 13, k and n), even though stimulation had ended and was no longer being delivered in the Wait Zone. No changes were observed in WZDM, or with prior stimulation for the two most preferred flavors or all flavors (Fig. 13, l-m and o-p). The persistent effect of Offer Zone stimulation during counterfactual re-evaluation in the Wait Zone, even after light delivery had ended, suggests that enhanced release of dopamine altered both current and future decision-making for OZDM, including both the factual (evaluation) and counterfactual (re-evaluation) components. Importantly, the absence of these effects in WZDM suggests that individual differences in decision-making strategy can dictate the manner in which dopamine dynamics regulate behavioral outcome.

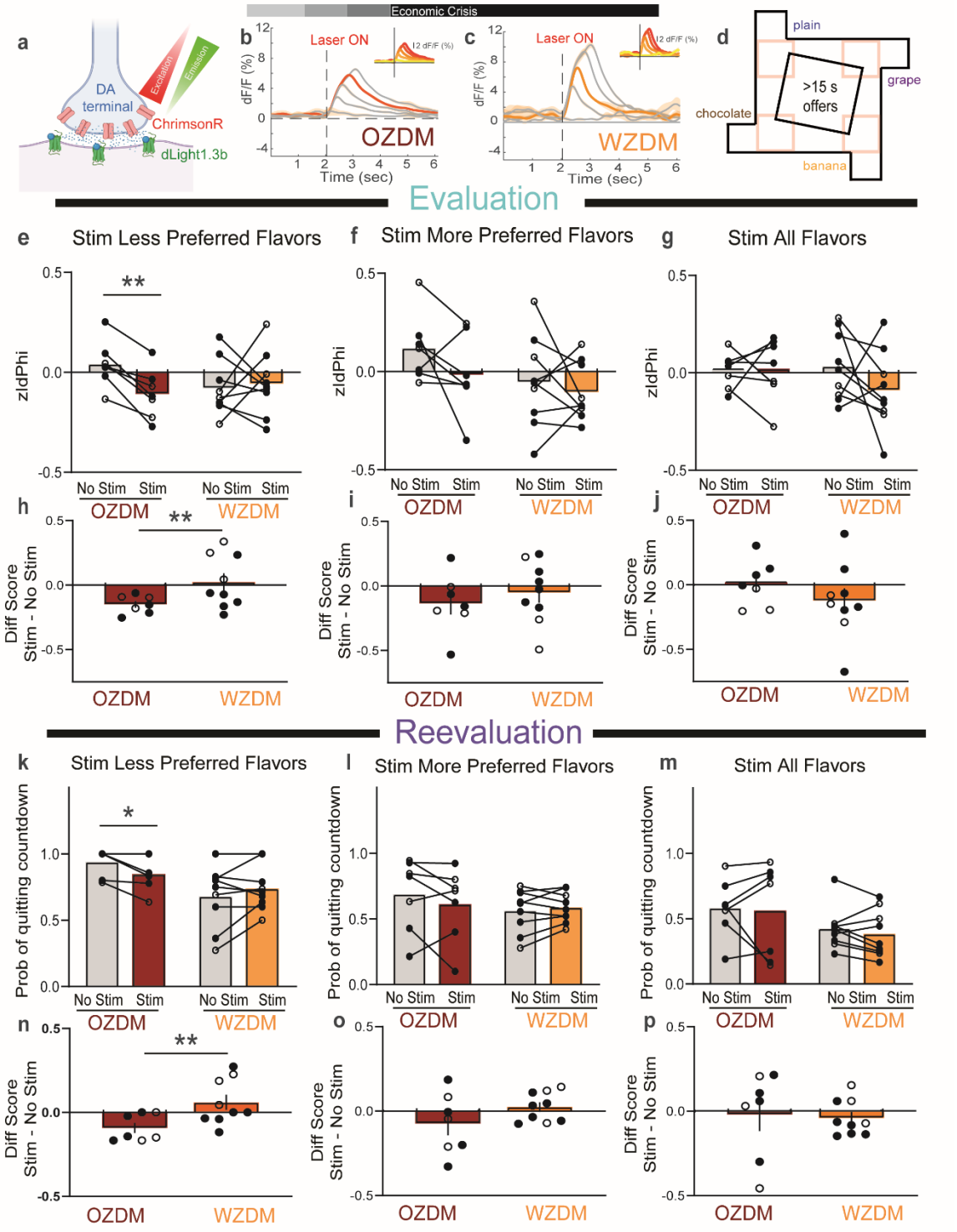

Figure 13 | Optogenetic enhancement of dopamine release has strategy-specific effects on decision evaluation and re-evaluation. (**a**) Schematic showing ChrimsonR expression in dopamine terminals, and dLight1.3b expression in NAc cells. (**b to c**) Dopamine response to different pulse numbers (0, 5, 10, 15, or 20; 589 nm, 20 Hz) in OZDM (b) and WZDM (c), highlighting calibrated stimulation parameters. (**d**) Offer Zone locations (red squares) where optogenetic stimulation was delivered on offers > 15 s. (**e to g**) zIdPhi on trials without stimulation (No Stim) and with stimulation (Stim) on less

preferred flavors (e), more preferred flavors (f), or all flavors (g). (**h to j**) Change in zIdPhi caused by stimulation on less preferred flavors (h), more preferred flavors (i), or all flavors (j). (**k to m**) Probability of quitting on No Stim and Stim trials for less preferred flavors (k), more preferred flavors (l), or all flavors (m). (**n to p**) Change in probability of quitting caused by stimulation on less preferred flavors (n), more preferred flavors (o), or all flavors (p). Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *$p<0.05$, **$p<0.01$ [ANOVA followed by Fisher's LSD post-hoc test]; see Data Table 1 for complete statistics.

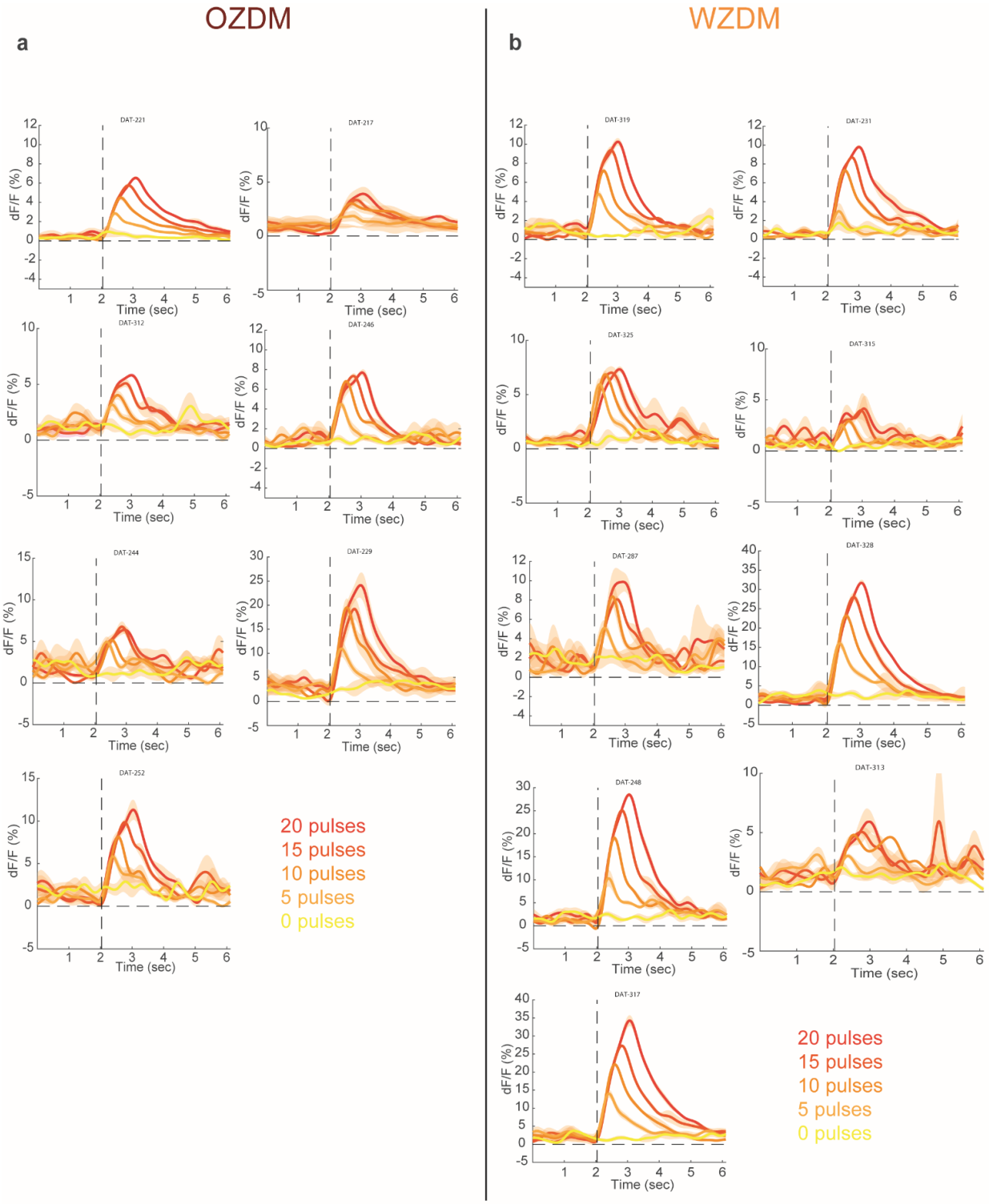

Figure 14 | Individually-tailored optogenetic stimulation responses. a) Optogenetic stimulation response curves for each OZDM and **b)** WZDM animals using the following parameters— pulse number: 0, 5, 10, 15, or 20; wavelength: 589 nm; frequency: 20 Hz.

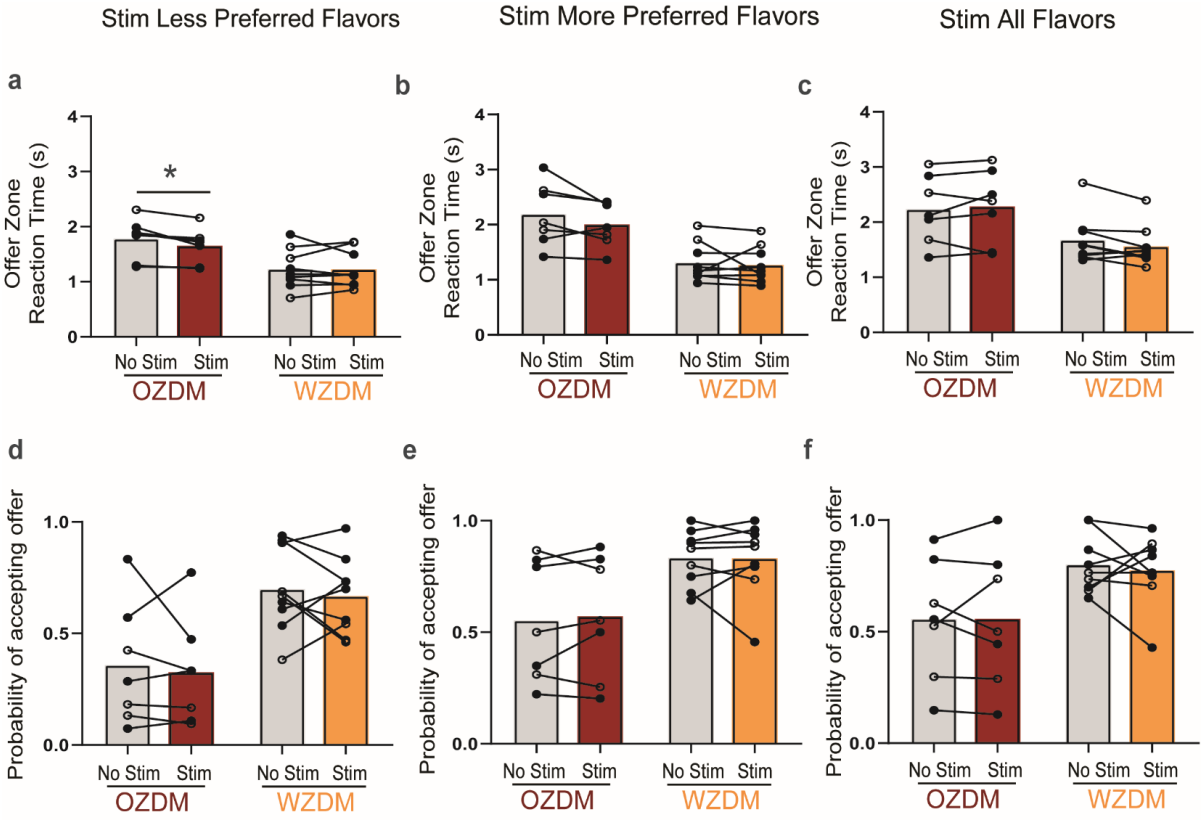

Figure 15 | Optogenetic stimulation effects on reaction time and offer acceptance. a) Reaction times during evaluation with stimulation in Less, **b)** More, and **c)** all preferred flavors. **d)** Probability of accepting offers on stimulated trials at Less, **e)** More, and **f)** all preferred flavors. Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *p<0.05, [ANOVA followed by Fisher's LSD post-hoc test]; see Data Table 1 for complete statistics.

Optogenetic inhibition of dopamine influenced the counterfactual re-evaluation of decisions

Since optogenetic stimulation reduced deliberative-like behaviors and decreased the probability of quitting, we wanted to test whether these effects would be bidirectionally modulated by optogenetic inhibition of dopamine terminals. To this end, we first validated that inhibition of terminals, in fact, reduced dopamine release in the NAc. In a separate group of mice expressing the inhibitory halorhodpsin in dopamine neurons, we delivered light (589 nm, 2 seconds continuous, 6-8mW) coincident with pellet delivery to inhibit the robust increases we observed during reward receipt. This manipulation significantly blocked post-Earn increases in dopamine across mice (Fig. 16a-d).

Using these parameters for photoinhibition, we again delivered light bilaterally in the Offer Zone on half of all offers >15 sec (Fig. 17a), on less preferred, more preferred, and all flavors, on separate days, as described earlier with optogenetic stimulation. To our surprise, none of these manipulations significantly altered deliberative behaviors in either decision-maker type (Fig. 17b-d). However, it is possible that inhibiting on offers >15 sec produces very minimal decreases in dopamine, since these bad offers already elicit only modest increases in dopamine endogenously (Fig 4g, i).

Nevertheless, the question still remained whether decreases in dopamine during re-evaluation causally contributed to quitting behaviors. To investigate this, we inhibited dopamine terminals during re-evaluation on accepted offers >15 sec at less preferred, more preferred, and all flavors, on separate days (Fig. 18a). We were unable to analyze the effects of inhibiting dopamine terminals at less preferred flavors, because animals were unlikely to accept offers >15 sec at these restaurants, which severely limited the number of trials available for analysis. However, we

found that photoinhibition during re-evaluation at the more preferred flavor increased the probability of quitting accepted offers during the countdown for both OZDM and WZDM (Fig. 18b,c). The lack of strategy-specific findings was consistent with our fiber photometry data demonstrating that both groups exhibited decreases in dopamine signaling prior to quitting (Fig. 4q, t), and suggest a causal role for dopamine dips in the quitting process. Recapitulating this decrease in dopamine sufficiently altered re-evaluation and promoted change-of-mind behaviors, which are linked to counterfactual processes.

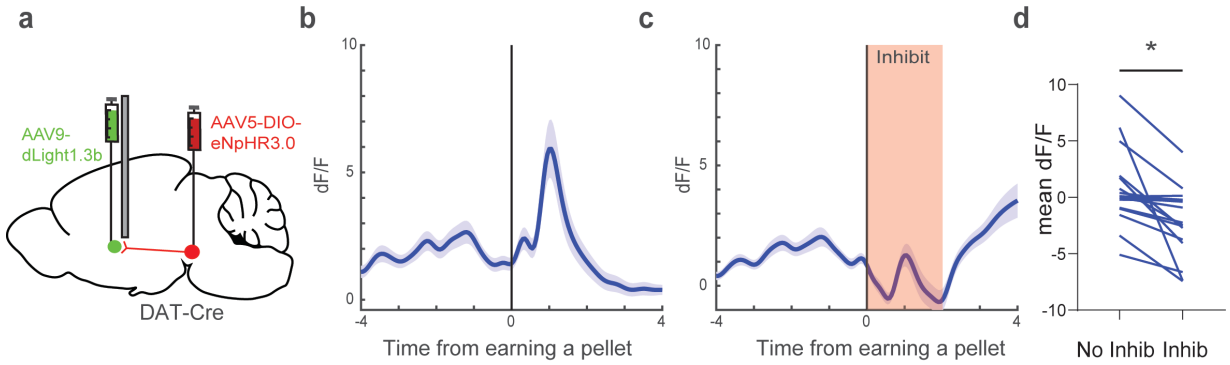

Figure 16 | Validation of optogenetic inhibition at terminals a) Schematic of viral strategy **b)** Dopamine response to earning a pellet with and **c)** without inhibition at terminals **d)** Average dopamine response with and without 2 sec of continuous inhibition Data are mean +/- SEM for all panels; *p<0.05, [ANOVA followed by Fisher's LSD post-hoc test]; see Data Table 1 for complete statistics.

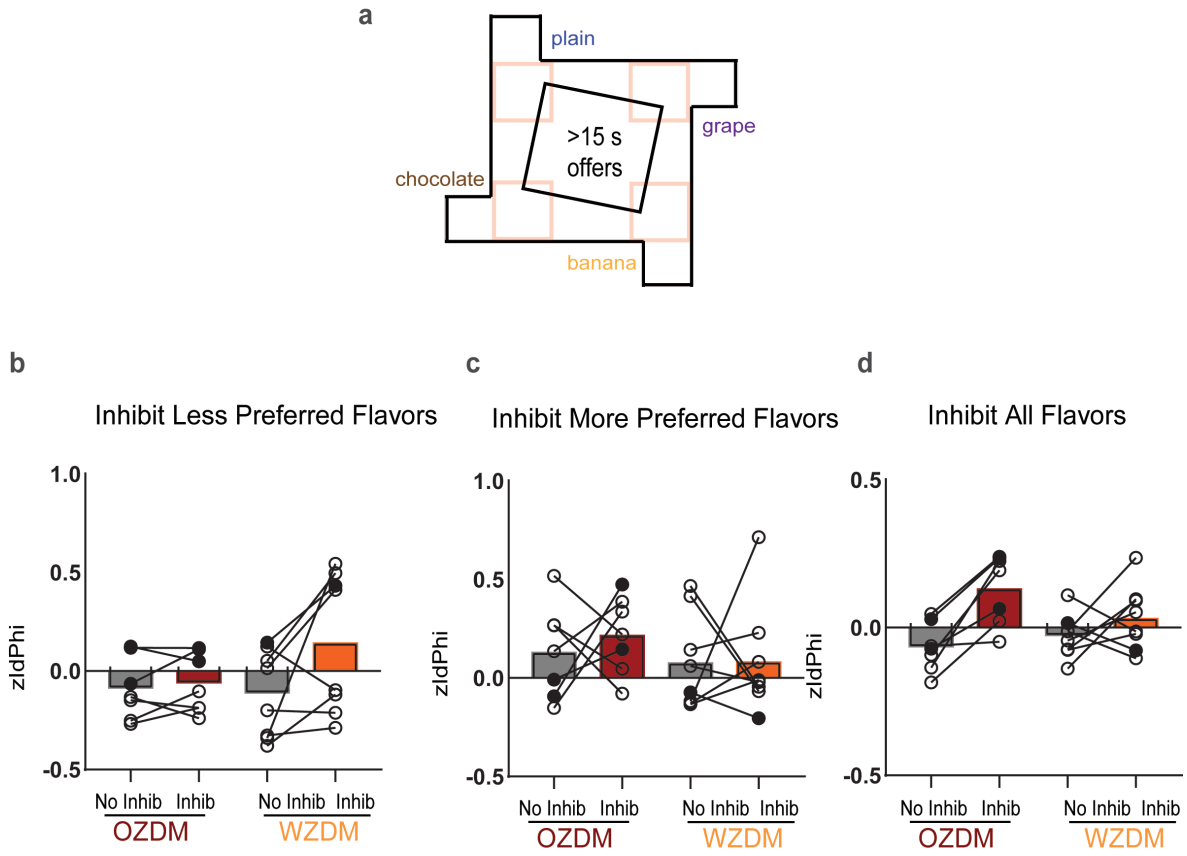

Figure 17 | Optogenetic inhibition of dopamine during evaluation does not significantly impact deliberative-like behaviors (a) Offer Zone locations (red squares) where optogenetic inhibition was delivered on offers > 15 s. (**b to d**) zIdPhi on trials without inhibition (No Inhib) and with inhibition (Inhib) on less preferred flavors (b), more preferred flavors (c), or all flavors (d). Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *p<0.05, [ANOVA]; see Data Table 1 for complete statistics.

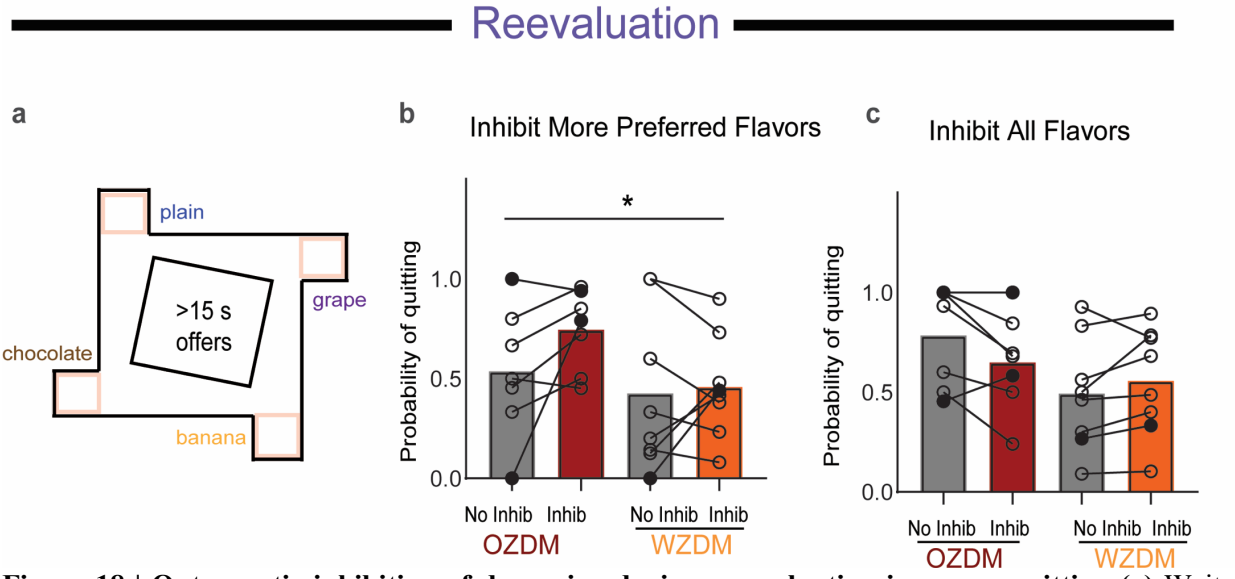

Figure 18 | Optogenetic inhibition of dopamine during re-evaluation increases quitting (a) Wait Zone locations (red squares) where optogenetic inhibition was delivered on offers > 15 s. (**b to c**) Probability of quitting on trials without inhibition (No Inhib) and with inhibition (Inhib) on more preferred flavors (b), or all flavors (c). Data are mean +/- SEM for all panels; open and filled circles represent female and male mice, respectively. *p<0.05, [ANOVA]; see Data Table 1 for complete statistics.

2.4 Discussion

Even when different behavioral strategies lead to the same outcome, the process to evaluate and re-evaluate decisions may diverge. We found that the process by which mice arrived at decisions was critical in shaping dopamine dynamics — individual differences in decision strategy and dopamine dynamics influenced the extent to which the mice expressed confidence in their options and engaged counterfactual processes to re-evaluate decisions. We identified two behavioral strategies with distinct relationships between dopamine and decision-making. Offer Zone Decision Makers (OZDM) engaged in more deliberative, future thinking strategies than Wait Zone Decision Makers (WZDM). Dopamine dynamics were strongest in the zone where animals made their decision. Importantly, dopamine dynamics reflected decision confidence during evaluation and future value during re-evaluation only in OZDM. This may suggest that animals with deliberative, future-thinking strategies have dopamine dynamics that impact confidence about future values. After physiologically calibrating our optogenetic stimulation parameters uniquely for each animal, we also established a causal contribution of dopamine to decision confidence in these mice with more deliberative, future-thinking phenotypes. This distinction in strategy may explain why OZDM were sensitive to optogenetic dopamine manipulations, whereas WZDM were unaffected. Furthermore, optogenetic stimulation yielded the greatest effects when stimulating at less preferred flavors, likely because behavior at less preferred flavors was more labile and amenable to alteration. Finally, optogenetically inhibiting dopamine terminals to mimic the dopamine dips we observed prior to quits, increased the probability of quitting across both decision maker types, causally linking dopamine decreases to counterfactual re-evaluation.

Dopamine dynamics have been widely investigated in the context of decision-making (Freels, Gabriel, Lester, & Simon, 2020; Le Heron et al., 2020), with some evidence of individual

differences related to stimulus-reward learning(Flagel et al., 2011). However, in these studies, decision-making is represented as a discrete event occurring at the moment at which an agent makes a choice. In contrast, decision-making is a continuous and iterative process, encompassing evaluation before and re-evaluation after the choice. The evaluation of options often involves deliberation to accumulate confidence and arrive at decisions, while re-evaluation of decisions often involves counterfactual reasoning to imagine alternate outcomes.

Temporal difference reinforcement learning (TDRL) reward-prediction error (RPE) has been the prevailing theory regarding dopamine as a regulator of learning and decision-making (Amo et al., 2022; Lerner, Holloway, & Seiler, 2021; Maes et al., 2020). While other theoretical proposals have been made, the behavioral paradigms used to test this theory are limited in their ability to capture important decision-making factors like confidence or counterfactual valuation, cognitive processes that underlie experiences of regret. Regret is distinct from disappointment — regret arises from mistakes of one's own agency, while disappointment reflects unexpected losses (Coricelli et al., 2005; Steiner & Redish, 2014; Zeelenberg & Pieters, 1999). RPE gives access to the latter through negative prediction errors, but dopamine's contributions to the former remain unstudied. More recent evidence highlighting the role of dopamine in signaling causal associations (Jeong et al., 2022), policies (Coddington et al., 2023), or perceived saliency (Kutlu et al., 2021) provide compelling evidence to challenge RPE, but do not assess these complex decision-making strategies that take into account counterfactual outcomes such as regret and re-evaluation. We found that dopamine signals in mice reflected that counterfactual information, in a manner related to the strategies individual animals used to achieve their goals.

Our findings regarding dopamine dips during quits also add to a literature that has largely focused on increases in dopamine levels, due to limitations in quantifying decreases in dopamine

with classic methods like fast-scan cyclic voltammetry. When decreases have been reported, they have been associated with external noxious stimuli (Goedhoop et al., 2022; Liu et al., 2022; Stelly et al., 2019) or externally-driven disappointment (Iino et al., 2020; Schultz et al., 1993, 1997; Tobler, Dickinson, & Schultz, 2003), but never to internally-driven cognitive events such as counterfactuals and re-evaluation. Our data indicate that dips and rebounds in dopamine can encode unique and distinct aspects of counterfactuals and enable self-directed re-evaluation. Further, we found that dopamine manipulations interacted with confidence, which took into account both actual and counterfactual outcomes. By unveiling distinct decision-making strategies, we find that mesolimbic dopamine conveys information about past and future value and scales with decision confidence, specifically for behavioral strategies that depend on computations related to future outcomes. These individual differences in the fundamental operation of mesolimbic dopamine could present unique individual vulnerabilities to dopamine dysfunction and associated neuropsychiatric conditions (Flagel et al., 2011; Saunders & Robinson, 2011; Schmack, Bosc, Ott, Sturgill, & Kepecs, 2021; Tye et al., 2013; Westbrook et al., 2020; Willmore, Cameron, Yang, Witten, & Falkner, 2022).

Chapter Three

Conclusion

Complex neuroeconomic models of decision-making

There is a pressing need to uncover the computations underlying decision-making in order to understand how animals interface with their environment. To this end, it is critical to model ethologically relevant, complex decision-making within our experimental confines.

Simple behavioral models may not be sufficient to capture nuanced decision-making dynamics since measurable behavioral outputs may be limited. Particularly in rodents, many behavioral tasks rely on port entry, anticipatory licking, or reaction time latency as a surrogate for learning or decision execution (Coddington et al., 2023)(Starkweather et al., 2017). While useful, these behavioral outputs do not have a wide dynamic range and therefore may mask subtleties in decision-making, especially in the context of competing options and values.

Moreover, artificially imposed "trials" and experimentally enforced choices create contrived settings which restrict our ability to probe natural decision processes. Decision-making is often conceptualized as a momentary event, occurring on a trial-by-trial basis, when in fact it is a more continuous and iterative process computationally.

The behavioral paradigm used for this thesis work leveraged neuroeconomic approach to interrogating decision-making. This work uncovered distinct decision-making phenotypes that relate to different economic strategies used to increase earnings. Furthermore, individualized flavor and waiting preferences allowed us to assess measures of past and future value for each animal. This work provides a sophisticated behavioral framework for accessing decision processes while considering unique economic strategies and valuation systems across individuals.

The work described in Chapter 2 leveraged a complex neuroeconomic decision-making task which is entirely self-driven. There were various features of this task that allowed us to probe the complexities of decision-making powerfully:

1) Because this is a foraging-based task, the initiation and termination of each trial was entirely volitional. The inter-trial-interval and inter-stimulus-interval transition states were not forced, and thus the quitting of trials was an *active* re-evaluation process rather than a simple passive omission of a response. This allowed us to probe not only simple metrics of value, but also more complex change-of-mind behaviors which are associated with counterfactual value processing.
2) Animals had the ability to both evaluate and re-evaluate their decisions at any point during the countdown through a moment-to-moment decision process. This allowed us to resolve decision-making on a more continuous timescale, rather than interpreting decision-making as a discrete event.
3) The complex nature of the task revealed nuanced differences in decision strategy. Exposing differences in decision computations is critical to grasping the heterogeneity of decision-

making and its underlying neural mechanisms. Distinct neural circuits and neural computations may underlie different valuation systems. Moreover, the same behavioral output may represent one computation in certain contexts, while representing another computation in another context, which is why elaborate behavioral paradigms are useful for disentangling distinct decision strategies.

4) Individual differences in flavor preferences and willingness to wait were calculated on a daily basis. This allowed for more accurate representations of value subjective to each animal, rather than imposing a single, common value state across all individuals. This also allows access to multiple dimensions of reward valuation like price, budget constraints, and opportunity cost.

Dopamine dynamics reflect decision strategy, counterfactual value, and decision confidence

In Chapter 2, we first uncovered distinct behavioral phenotypes which reflected different economic strategies to increase earnings over time. One group, Offer Zone Decision Makers (OZDM), made initial Accept and Skip decisions according to offer delay and developed strategies over training to only accept offer delays for which they were subsequently willing to wait. In contrast, the Wait Zone Decision Makers (WZDM), maintained a strategy of accepting most offers irrespective of the delay, and then deciding whether to commit or quit during re-evaluation.

It is not surprising that there may exist different strategic approaches to maximize earnings, but our findings raise interesting questions about the various ways in which animals may solve the problem of limited resources. Perhaps some decision-making types place a greater emphasis on

the future consequences of making a decision, but once the decision is made, the individual is fully committed. Others may be less committal about the future and continue to consider their decision even after it is made. Furthermore, decision-makers may vary in the degree to which they deliberate their decisions about the future.

We next wanted to determine whether the different behavioral phenotypes mapped onto differences in mesolimbic dopamine dynamics between groups. Specific differences in dopamine dynamics during evaluation and re-evaluation corresponded with differences in decision strategy. Most interestingly, I found that dips in dopamine preceding quitting in both decision maker types. As described in detail earlier, decreases in dopamine are predicted by reinforcement learning models, yet there are few publications that report dopamine reductions in experimental settings. In cases where decreases have been reported, they have largely been associated with aversive stimuli or reward omissions. (These reports have been puzzling, with mixed evidence of both increases and decreases in dopamine when rewards are omitted (Ishino et al., 2023), and reports of increases in dopamine to aversive stimuli during negative reinforcement (Kutlu et al., 2021)(Diao et al., 2021)). An important distinction with this work is that there were no aversive stimuli or specific cues informing the animal when or whether to quit. Instead, decreases observed during quitting were associated with an entirely self-driven, volitional process as opposed to being externally imposed (e.g., reward omission).

At its core, quitting can be thought of as the culmination of a re-evaluation process. This is because quitting is a change-of-mind process, since the animal has already accepted the offer and the default during the countdown is to commit. The animal must *opt in* to quitting, because inaction

during the countdown defaults to earning a pellet. In order to exhibit a change-of-mind, the agent must presumably consider alternative options, possibly imagining alternate pasts and possible futures until ultimately deciding to quit. These are counterfactual representations of what might have been or what could possibly be.

We found that dopamine dipped before and rapidly recovered after quitting. Interestingly, the dip in dopamine that occurred pre-quit period was associated with future value, while the increase in dopamine that occurred post-quit period was associated with past value. This suggests that distinct portions of an event-aligned dopamine signal can pertain to distinct computations. For example, it is possible that the dip that occurred pre-quit is related to regret-like counterfactuals, while the increase post-quit is related to relief-like counterfactuals. These findings raise the interesting possibility that event-locked dopamine signals may warrant further exploration as the pre- and post-event dynamics may correlate with different cognitive processes. Furthermore, the evidence that dopamine may reflect past and future value bolsters the point that the TDRL model should be expanded to include counterfactual prediction terms. In addition to accounting for actual and expected outcomes, the model may benefit from integrating counterfactual predictions about what could have been.

Dips in dopamine before quits, though present in both decision-maker types, specifically reflected favorable future values only in Offer Zone Decision Makers. This may be because the Offer Zone Decision Makers were more intently considering the future consequences of their actions. Further evidence for this was observed during evaluation when Offer Zone Decision Makers showed greater levels of deliberative-like behaviors and dopamine dynamics that related to the degree of

deliberation. This is a sign that Offer Zone Decision Makers engaged in greater future planning whereas Wait Zone Decision Makers may have displayed more snap-judgments. Prior work exploring deliberative-like behaviors in rodents in operant tasks has relied on the period before action selection as a surrogate for deliberation(Bercovici, Princz-Lebel, Tse, Moorman, & Floresco, 2018)(Vandaele, Lenoir, Vouillac-Mendoza, Guillem, & Ahmed, 2021), however this is an ambiguous metric since there is no specific behavioral correlate of deliberation in these cases. While time before action selection can include deliberation, it can also be obfuscated by other confounding cognitive processes as well as motivational drive. Therefore, it was important for us to utilize VTE as a reliable behavioral measure of deliberative-like behaviors in rodents. Using VTE we were able to contrast dopamine dynamics based on high- and low- deliberative states.

After monitoring dopamine dynamics during decision-making we conducted causal manipulations during evaluation and re-evaluation. First, optogenetic enhancement of dopamine on unfavorable offers (long delays at least preferred flavors) reduced deliberative behavior in a strategy specific manner, only influencing Offer Zone Decision Makers who tended to be more future planning. What is more, the effects of dopamine augmentation during evaluation carried over into re-evaluation, after the stimulation had terminated, reducing the probability of quitting. In summary, dopamine enhancement on bad offers reduced deliberation and increased the propensity to commit to an offer the animal would have otherwise likely quit. This finding suggests that evaluation and re-evaluation are inextricably linked. However, questions still remain about whether the dips in dopamine seen during re-evaluation were causally linked to quitting behaviors. To this end, I inhibited dopamine terminals during the countdown in re-evaluation. Interestingly, inhibiting terminals increased the probability of quitting at more preferred flavors across both decision-maker

types. The effect in both decision-maker types aligns with the decreases in dopamine measured in both groups with fiber photometry. This is some of the first evidence that projection-specific, terminal inhibition of dopamine release is causally related to a counterfactual process. Importantly, in conducting our manipulations, we ensured that our stimulation parameters evoked or inhibited dopamine release in a manner that paralleled endogenous dynamics, so as to ensure that none of our manipulations were supraphysiologic.

Collectively, this work emphasizes the importance of dissecting behavioral phenotypes when assessing behavior and neural dynamics. Behavioral subtypes can represent distinct decision-making processes that may be obscured or masked when grouped together. Particularly within the dopamine field, it is critical to unmask these strategic differences because they may relate to individual differences in value processing, decision confidence, expectations and counterfactual processes. These decisional differences may contribute to the tremendous variation in dopamine theories.

In turn, the findings in this thesis could be used to inform models of reinforcement learning. Specifically, I argue that TDRL algorithms may benefit from expanding to include counterfactual concepts within their framework. As mentioned earlier, the foundational TDRL algorithms calculate value dependent on the actual state of the world. This is learned by computing the difference between actual and predicted changes in value at each state. While this takes into account expectations and actual values, there is no term in the model that considers counterfactual values. In cases where there are limited resources, competing options, or conflicting motivations, the value of each state may depend on *both* actual and counterfactual representations. Representing

this in more modern conceptions of TDRL may account for discrepancies that original TDRL learning cannot explain.

Chapter Four

Implications and Future Directions

Implications for computational psychiatry

Sophisticated behavioral models of decision-making may engage neural computations that can 1) otherwise be obscured by simpler models, and 2) translate more directly to humans. Uncovering the neural correlates of these decision processes allows us to contend with decision-making dysregulations seen in psychiatric disease. Simply put, in order to understand what happens in the brain when things go wrong, we must clearly understand what happens when things go right.

It is true that deficits in decision-making span various psychiatric diseases. Psychiatric disorders often result in sub-optimal performance in reward-seeking and loss-avoidance decision tasks (Suzuki, Yamashita, & Katahira, 2021). Patients with depression often have reduced physical and cognitive effort discounting (Hershenberg et al., 2016) and are more emotionally reactive to unfair decision outcomes (Harlé, Allen, & Sanfey, 2010). Patients with schizophrenia regularly show impairments in weighing risks and reward when faced with conflicting options (Yip, Sacco, George, & Potenza, 2009). Eating disorders also relate to bidirectional decisional deficits—while binge eating disorder is correlated with steeper delay-discounting (preferring sooner rewards) (Brassard & Balodis, 2021), anorexia-nervosa is correlated with shallower delay-discounting (preferring later rewards) (Amlung et al., 2019).

However, while many aforementioned psychiatric disorders can manifest in decision-making aberrancies, there is great heterogeneity in the underlying systems that go awry. Distinct valuation systems may contribute to similar decision-making phenotypes in vastly different disorders. Ostensibly similar deficits could stem from entirely different network dysregulations within the brain. This has plagued the neuroscience field for decades as we try to resolve the computational processes that breakdown in depression versus schizophrenia versus obsessive compulsive disorder, etc.

It is difficult to disentangle all of the contributory behavioral micro-changes that ultimately alter decision-making in psychiatric illness. Often, we only see the final, gross output (e.g., steeper delay discounting, or greater impulsivity, or altered risk-taking), yet the underlying decisional computations remain elusive (e.g., subjective valuation, deliberation, counterfactual reasoning). Our ability to resolve the neural circuits and networks underlying diverse psychiatric disorders relies on our power to disambiguate these decision-making processes.

Complex translational models of decision-making that span across species can be useful diagnostic and treatment tools in psychiatry. As I previously described, the behavioral task described in chapter 2 gives us access to more in-depth metrics of value during decision-making. While our experiments were conducted in mice, the task is translatable across species like rats, nonhuman-primates, and humans (Redish et al., 2021)(Abram, Breton, Schmidt, Redish, & MacDonald, 2016)(Sweis, Abram, et al., 2018). This allows for bidirectional experimentation whereby behavioral findings in humans with psychiatric disease can be used to inform behavioral studies

in rodents, and circuit-specific findings in rodents can be used to inform the neural underpinnings of decision-making and psychiatric disease in humans.

Implications for reinforcement learning models of dopamine

Dopamine is a critical neural substrate for learning and decision-making. As hypothesized by reinforcement learning theory, dopamine responses have reported errors in reward predictions in a way that remarkably resembles error signals in the TDRL algorithm. While the RPE framework has predominated as the most prevalent model, there is strong impetus to address its limits in explaining many perplexing dopamine findings in the decision-making literature.

Recently, many have proposed alternate theories that dopamine reflects the inferred cause of outcomes retrospectively (Jeong et al., 2022), the perceived saliency of stimuli (Kutlu et al., 2021), or the policy learning rate (Coddington et al., 2023). Although these theories challenge the field with compelling evidence, the experiments all came from a very restricted behavioral landscape which primarily relies on simple behavioral outputs like cued approach, cued avoidance, immobility/freezing, and anticipatory licking in a head-fixed preparation. In this thesis, I argue that these behavioral contexts may not be sufficient to access facets of decision-making that could inform our models of learning. Many of these behaviors are binary outputs (e.g., lick or no lick, freeze or no freeze), and the actions selected in these conditions are distinct from actions selected when there are conflicting choices and competing values. Specifically, none of these aforementioned behaviors can expose the deliberative or counterfactual considerations of the agent. Furthermore, neither the canonical TDRL nor the more recently proposed ANCCR, KCS,

ACTR algorithms account for counterfactual value in their models. This is crucial because so much of decision-making requires integrating the actual with the counterfactual.

Future directions

The work conducted in this dissertation helps illuminate the role of dopamine in a complex economic environment. However, as is often the case, these findings raise additional questions to be answered, and lay the groundwork for further studies.

The experiments described in this thesis focused on neurochemical measurements of dopamine. However, it is still unknown how postsynaptic neurons within the NAc respond to the observed fluctuations in dopamine in this decision-making context. MSNs within the NAc are known to integrate dopaminergic signals from the VTA with glutamatergic signals from the cortex and thalamus (Wickens, Horvitz, Costa, & Killcross, 2007)(Doig, Moss, & Bolam, 2010)(Andrianarivelo, Saint-Jour, Walle, Trifilieff, & Vanhoutte, 2019)(F. J. S. Lee et al., 2002), but it is unknown how decreases in dopamine are computationally registered by downstream D1- and D2-expressing MSNs. Along these lines, we know that dopamine release and dopamine neuron firing are separable (Mohebi et al., 2019). This raises the question about whether any decreases seen in dopamine, during re-evaluation or otherwise, are correlated with decreased dopamine neuron activity or are locally controlled by interneurons at the terminal region.

While midbrain dopamine neurons project most densely to striatal regions, there is also significant dopaminergic innervation of prefrontal cortical regions. We know that the orbitofrontal cortex,

specifically, is critical for many of the cognitive aspects of decision-making, like learning switched contingencies in reversal learning (Izquierdo, Suda, & Murray, 2004), alternating between options (Mishkin, Vest, Waxler, & Rosvold, 1969), and extinction learning(Butter, Mishkin, & Rosvold, 1963). Orbitofrontal regions have been known to represent cognitive maps of state spaces (Schuck, Cai, Wilson, & Niv, 2016) and mental simulations of possible outcomes during decision-making (Gallagher, McMahan, & Schoenbaum, 1999). However, it is unknown to what extent dopamine dynamics track counterfactual values in orbitofrontal regions. The work in this dissertation can serve as a template to advance the field using state-of-the-art technology to investigate these questions.

Bibliography

Abe, H., & Lee, D. (2011). Distributed Coding of Actual and Hypothetical Outcomes in the Orbital and Dorsolateral Prefrontal Cortex. *Neuron*. https://doi.org/10.1016/j.neuron.2011.03.026

Abercrombie, E. D., Keefe, K. A., DiFrischia, D. S., & Zigmond, M. J. (1989). Differential Effect of Stress on In Vivo Dopamine Release in Striatum, Nucleus Accumbens, and Medial Frontal Cortex. *Journal of Neurochemistry*. https://doi.org/10.1111/j.1471-4159.1989.tb09224.x

Abraham, C., & Sheeran, P. (2004). Deciding to exercise: The role of anticipated regret. *British Journal of Health Psychology*. https://doi.org/10.1348/135910704773891096

Abram, S. V., Breton, Y. A., Schmidt, B., Redish, A. D., & MacDonald, A. W. (2016). The Web-Surf Task: A translational model of human decision-making. *Cognitive, Affective and Behavioral Neuroscience*. https://doi.org/10.3758/s13415-015-0379-y

Aitken, T. J., Greenfield, V. Y., & Wassum, K. M. (2016). Nucleus accumbens core dopamine signaling tracks the need-based motivational value of food-paired cues. *Journal of Neurochemistry*, *136*(5), 1026–1036. https://doi.org/10.1111/jnc.13494

Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., … McCabe, R. E. (2019). Delay Discounting as a Transdiagnostic Process in Psychiatric Disorders: A Meta-analysis. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2019.2102

Amo, R., Matias, S., Yamanaka, A., Tanaka, K. F., Uchida, N., & Watabe-Uchida, M. (2022). A gradual temporal shift of dopamine responses mirrors the progression of temporal difference error in machine learning. *Nature Neuroscience*, *25*(8), 1082–1092.

https://doi.org/10.1038/s41593-022-01109-2

Andrianarivelo, A., Saint-Jour, E., Walle, R., Trifilieff, P., & Vanhoutte, P. (2019). Modulation and functions of dopamine receptor heteromers in drugs of abuse-induced adaptations. *Neuropharmacology*. https://doi.org/10.1016/j.neuropharm.2018.12.003

Bäckman, C. M., Malik, N., Zhang, Y. J., Shan, L., Grinberg, A., Hoffer, B. J., … Tomac, A. C. (2006). Characterization of a mouse strain expressing Cre recombinase from the 3′ untranslated region of the dopamine transporter locus. *Genesis (United States)*, *44*(8), 383–390. https://doi.org/10.1002/dvg.20228

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*. https://doi.org/10.1016/0010-0277(94)90018-3

Bercovici, D. A., Princz-Lebel, O., Tse, M. T., Moorman, D. E., & Floresco, S. B. (2018). Optogenetic dissection of temporal dynamics of amygdala-striatal interplay during risk/reward decision making. *ENeuro*. https://doi.org/10.1523/ENEURO.0422-18.2018

Berridge, K. C., & Robinson, T. E. (1998, December). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, Vol. 28, pp. 309–369. https://doi.org/10.1016/S0165-0173(98)00019-8

Bertler, Å., & Rosengren, E. (1959). Occurrence and distribution of dopamine in brain and other tissues. *Experientia*. https://doi.org/10.1007/BF02157069

Boldt, A., Schiffer, A. M., Waszak, F., & Yeung, N. (2019). Confidence Predictions Affect Performance Confidence and Neural Preparation in Perceptual Decision Making. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-40681-9

Boulanger-Bertolus, J., Parrot, S., Doyère, V., & Mouly, A. M. (2021). Dorsal striatum and the

temporal expectancy of an aversive event in Pavlovian odor fear learning. *Neurobiology of Learning and Memory*. https://doi.org/10.1016/j.nlm.2021.107446

Brassard, S. L., & Balodis, I. M. (2021). A review of effort-based decision-making in eating and weight disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. https://doi.org/10.1016/j.pnpbp.2021.110333

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.36.8.917

Brischoux, F., Chakraborty, S., Brierley, D. I., & Ungless, M. A. (2009). Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0811507106

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. *Neuron*. https://doi.org/10.1016/j.neuron.2010.11.022

Butter, C. M., Mishkin, M., & Rosvold, H. E. (1963). Conditioning and extinction of a food-rewarded response after selective ablations of frontal cortex in rhesus monkeys. *Experimental Neurology*. https://doi.org/10.1016/0014-4886(63)90094-3

Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the Suppression of Inferences. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1998.2622

Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*. https://doi.org/10.1257/0022051053737843

Cameron, D. L., & Williams, J. T. (1993). Dopamine D1 receptors facilitate transmitter release. *Nature*. https://doi.org/10.1038/366344a0

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J. R., & Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*. https://doi.org/10.1126/science.1094550

Carlsson, A., Lindqvist, M., & Magnusson, T. (1957). 3,4-Dihydroxyphenylalanine and 5-hydroxytryptophan as reserpine antagonists [16]. *Nature*. https://doi.org/10.1038/1801200a0

Carlsson, A., Lindqvist, M., Magnusson, T., & Waldeck, B. (1958). On the presence of 3-hydroxytyramine in brain. *Science*. https://doi.org/10.1126/science.127.3296.471

Chan, C. S., Gertler, T. S., & Surmeier, D. J. (2010). A molecular basis for the increased vulnerability of substantia nigra dopamine neurons in aging and Parkinson's disease. *Movement Disorders*. https://doi.org/10.1002/mds.22801

Chang, C. Y., Gardner, M., Di Tillio, M. G., & Schoenbaum, G. (2017). Optogenetic Blockade of Dopamine Transients Prevents Learning Induced by Changes in Reward Features. *Current Biology*. https://doi.org/10.1016/j.cub.2017.09.049

Choi, W. Y., Balsam, P. D., & Horvitz, J. C. (2005). Extended habit training reduces dopamine mediation of appetitive response expression. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1498-05.2005

Coddington, L. T., Lindo, S. E., & Dudman, J. T. (2023). Mesolimbic dopamine adapts the rate of learning from action. *Nature*, *614*(7947), 294–302. https://doi.org/10.1038/s41586-022-05614-z

Collins, A. L., Greenfield, V. Y., Bye, J. K., Linker, K. E., Wang, A. S., & Wassum, K. M. (2016). Dynamic mesolimbic dopamine signaling during action sequence learning and

expectation violation. *Scientific Reports*. https://doi.org/10.1038/srep20231

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience*, *8*(9), 1255–1262. https://doi.org/10.1038/nn1514

Dabney, W., Kurth-Nelson, Z., Uchida, N., Starkweather, C. K., Hassabis, D., Munos, R., & Botvinick, M. (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature*. https://doi.org/10.1038/s41586-019-1924-6

de Jong, J. W., Afjei, S. A., Pollak Dorocic, I., Peck, J. R., Liu, C., Kim, C. K., … Lammel, S. (2019). A Neural Circuit Mechanism for Encoding Aversive Stimuli in the Mesolimbic Dopamine System. *Neuron*. https://doi.org/10.1016/j.neuron.2018.11.005

Diao, Z., Yao, L., Cheng, Q., Wu, M., Di, Y., Qian, Z., … Ren, W. (2021). Involvement of Midbrain Dopamine Neuron Activity in Negative Reinforcement Learning in Mice. *Molecular Neurobiology*. https://doi.org/10.1007/s12035-021-02515-6

Doig, N. M., Moss, J., & Bolam, J. P. (2010). Cortical and thalamic innervation of direct and indirect pathway medium-sized spiny neurons in mouse striatum. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1623-10.2010

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*. https://doi.org/10.1126/science.1077349

Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., … Akil, H. (2011). A selective role for dopamine in stimulus-reward learning. *Nature*, *469*(7328), 53–59. https://doi.org/10.1038/nature09588

Freels, T. G., Gabriel, D. B. K., Lester, D. B., & Simon, N. W. (2020). Risky decision-making predicts dopamine release dynamics in nucleus accumbens shell.

Neuropsychopharmacology, *45*(2), 266–275. https://doi.org/10.1038/s41386-019-0527-0

Frydman, C., & Camerer, C. (2016). Neural evidence of regret and its implications for investor behavior. *Review of Financial Studies*, *29*(11), 3108–3139. https://doi.org/10.1093/rfs/hhw010

Fulford, A. J., & Marsden, C. A. (2007). An intact dopaminergic system is required for context-conditioned release of 5-HT in the nucleus accumbens of postweaning isolation-reared rats. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2007.08.002

Gallagher, M., McMahan, R. W., & Schoenbaum, G. (1999). Orbitofrontal cortex and representation of incentive value in associative learning. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.19-15-06610.1999

Gardner, M. P. H., Schoenbaum, G., & Gershman, S. J. (2018). Rethinking dopamine as generalized prediction error. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2018.1645

German, C. L., Baladi, M. G., McFadden, L. M., Hanson, G. R., & Fleckenstein, A. E. (2015). Regulation of the dopamine and vesicular monoamine transporters: Pharmacological targets and implications for disease. *Pharmacological Reviews*. https://doi.org/10.1124/pr.114.010397

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science*. https://doi.org/10.1111/1467-7687.00309

Gilbert, A. N., Fridlund, A. J., & Sabini, J. (1987). Hedonic and social determinants of facial displays to odors. *Chemical Senses*. https://doi.org/10.1093/chemse/12.2.355

Gilovich, T., & Medvec, V. H. (1995). The Experience of Regret: What, When, and Why. *Psychological Review*. https://doi.org/10.1037/0033-295X.102.2.379

Glickstein, M., & Stein, J. (1991). Paradoxical movement in Parkinson's disease. *Trends in Neurosciences*. https://doi.org/10.1016/0166-2236(91)90055-Y

Goedhoop, J. N., van den Boom, B. J. G., Robke, R., Veen, F., Fellinger, L., van Elzelingen, W., … Willuhn, I. (2022). Nucleus accumbens dopamine tracks aversive stimulus duration and prediction but not value or prediction error. *ELife*, *11*. https://doi.org/10.7554/eLife.82711

Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain*. https://doi.org/10.1093/brain/120.10.1805

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., … Berke, J. D. (2015). Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, *19*(1), 117–126. https://doi.org/10.1038/nn.4173

Harlé, K. M., Allen, J. J. B., & Sanfey, A. G. (2010). The impact of depression on social economic decision making. *Journal of Abnormal Psychology*. https://doi.org/10.1037/a0018612

Hart, W. E., Goldbaum, M., Côté, B., Kube, P., & Nelson, M. R. (1999). Measurement and classification of retinal vascular tortuosity. *International Journal of Medical Informatics*, *53*(2–3), 239–252. https://doi.org/10.1016/S1386-5056(98)00163-4

Hershenberg, R., Satterthwaite, T. D., Daldal, A., Katchmar, N., Moore, T. M., Kable, J. W., & Wolf, D. H. (2016). Diminished effort on a progressive ratio task in both unipolar and bipolar depression. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2016.02.003

Hnasko, T. S., Chuhma, N., Zhang, H., Goh, G. Y., Sulzer, D., Palmiter, R. D., … Edwards, R. H. (2010). Vesicular Glutamate Transport Promotes Dopamine Storage and Glutamate Corelease In Vivo. *Neuron*. https://doi.org/10.1016/j.neuron.2010.02.012

Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*. https://doi.org/10.1016/S0306-4522(00)00019-1

Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips, P. E. M., & Graybiel, A. M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature*, *500*(7464), 575–579. https://doi.org/10.1038/nature12475

Iino, Y., Sawada, T., Yamaguchi, K., Tajiri, M., Ishii, S., Kasai, H., & Yagishita, S. (2020). Dopamine D2 receptors in discrimination learning and spine enlargement. *Nature*, *579*(7800), 555–560. https://doi.org/10.1038/s41586-020-2115-1

Ishino, S., Kamada, T., Sarpong, G. A., Kitano, J., Tsukasa, R., Mukohira, H., … Ogawa, M. (2023). Dopamine error signal to actively cope with lack of expected reward. *Science Advances*. https://doi.org/10.1126/sciadv.ade5420

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1921-04.2004

Jeong, H., Taylor, A., Floeder, J. R., Lohmann, M., Mihalas, S., Wu, B., … Namboodiri, V. M. K. (2022). Mesolimbic dopamine release conveys causal associations. *Science*, *378*(6626). https://doi.org/10.1126/science.abq6740

Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3761-07.2007

Josephs, R. A., Larrick, R. P., Steele, C. M., & Nisbett, R. E. (1992). Protecting the Self From the Negative Consequences of Risky Decisions. *Journal of Personality and Social Psychology*, *62*(1), 26–37. https://doi.org/10.1037/0022-3514.62.1.26

Kalenscher, T., & van Wingerden, M. (2011). Why we should use animals to study economic decision making - A perspective. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2011.00082

Kalhan, S., Redish, A. D., Hester, R., & Garrido, M. I. (2021). A salience misattribution model for addictive-like behaviors. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2021.02.039

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. *Punishment and Aversive Behavior*.

Kay, K., Chung, J. E., Sosa, M., Schor, J. S., Karlsson, M. P., Larkin, M. C., … Frank, L. M. (2020). Constant Sub-second Cycling between Representations of Possible Futures in the Hippocampus. *Cell*, *180*(3), 552-567.e25. https://doi.org/10.1016/j.cell.2020.01.014

Kaźmierczak, M., & Nicola, S. M. (2022). The Arousal-motor Hypothesis of Dopamine Function: Evidence that Dopamine Facilitates Reward Seeking in Part by Maintaining Arousal. *Neuroscience*. https://doi.org/10.1016/j.neuroscience.2022.07.008

Kim, H. G. R., Malik, A. N., Mikhael, J. G., Bech, P., Tsutsui-Kimura, I., Sun, F., … Uchida, N. (2020). A Unified Framework for Dopamine Signals across Timescales. *Cell*. https://doi.org/10.1016/j.cell.2020.11.013

Kishida, K. T., Saez, I., Lohrenz, T., Witcher, M. R., Laxton, A. W., Tatter, S. B., … Montague, P. R. (2016). Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 200–205. https://doi.org/10.1073/pnas.1513619112

Klapoetke, N. C., Murata, Y., Kim, S. S., Pulver, S. R., Birdsey-Benson, A., Cho, Y. K., …

Boyden, E. S. (2014). Independent optical excitation of distinct neural populations. *Nature Methods*, *11*(3), 338–346. https://doi.org/10.1038/nmeth.2836

Kurth-Nelson, Z., & Redish, A. D. (2009). Temporal-difference reinforcement learning with distributed representations. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0007362

Kutlu, M. G., Tat, J., Christensen, B. A., Zachry, J. E., & Calipari, E. S. (2023). Dopamine release at the time of a predicted aversive outcome causally controls the trajectory and expression of conditioned behavior. *Cell Reports*. https://doi.org/10.1016/j.celrep.2023.112948

Kutlu, M. G., Zachry, J. E., Melugin, P. R., Cajigas, S. A., Chevee, M. F., Kelly, S. J., … Calipari, E. S. (2021). Dopamine release in the nucleus accumbens core signals perceived saliency. *Current Biology*, *31*(21), 4748-4761.e8. https://doi.org/10.1016/j.cub.2021.08.052

Labouèbe, G., Lomazzi, M., Cruz, H. G., Creton, C., Luján, R., Li, M., … Lüscher, C. (2007). RGS2 modulates coupling between GABAB receptors and GIRK channels in dopamine neurons of the ventral tegmental area. *Nature Neuroscience*. https://doi.org/10.1038/nn2006

Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B., & Roeper, J. (2008). Unique Properties of Mesoprefrontal Neurons within a Dual Mesocorticolimbic Dopamine System. *Neuron*. https://doi.org/10.1016/j.neuron.2008.01.022

Lammel, S., Lim, B. K., & Malenka, R. C. (2014, January 1). Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology*, Vol. 76, pp. 351–359. https://doi.org/10.1016/j.neuropharm.2013.03.019

Larquet, M., Coricelli, G., Opolczynski, G., & Thibaut, F. (2010). Impaired decision making in schizophrenia and orbitofrontal cortex lesion patients. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2009.11.010

Laurent, V., & Balleine, B. W. (2015). Factual and counterfactual action-outcome mappings control choice between goal-directed actions in rats. *Current Biology*. https://doi.org/10.1016/j.cub.2015.02.044

Le Heron, C., Kolling, N., Plant, O., Kienast, A., Janska, R., Ang, Y. S., … Apps, M. A. J. (2020). Dopamine Modulates Dynamic Decision-Making during Foraging. *Journal of Neuroscience*, *40*(27), 5273–5282. https://doi.org/10.1523/JNEUROSCI.2586-19.2020

Lee, F. J. S., Xue, S., Pei, L., Vukusic, B., Chéry, N., Wang, Y., … Liu, F. (2002). Dual regulation of NMDA receptor functions by direct protein-protein interactions with the dopamine D1 receptor. *Cell*. https://doi.org/10.1016/S0092-8674(02)00962-5

Lee, R. S., Mattar, M. G., Parker, N. F., Witten, I. B., & Daw, N. D. (2019). Reward prediction error does not explain movement selectivity in dms-projecting dopamine neurons. *ELife*. https://doi.org/10.7554/eLife.42992

Lefevre, E. M., Pisansky, M. T., Toddes, C., Baruffaldi, F., Pravetoni, M., Tian, L., … Rothwell, P. E. (2020). Interruption of continuous opioid exposure exacerbates drug-evoked adaptations in the mesolimbic dopamine system. *Neuropsychopharmacology*, *45*(11), 1781–1792. https://doi.org/10.1038/s41386-020-0643-x

Lerner, T. N., Holloway, A. L., & Seiler, J. L. (2021, April 1). Dopamine, Updated: Reward Prediction Error and Beyond. *Current Opinion in Neurobiology*, Vol. 67, pp. 123–130. https://doi.org/10.1016/j.conb.2020.10.012

Levey, A. I., Hersch, S. M., Rye, D. B., Sunahara, R. K., Niznik, H. B., Kitt, C. A., … Ciliax, B. J. (1993). Localization of D1 and D2 dopamine receptors in brain with subtype-specific antibodies. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.90.19.8861

Liebenow, B., Jiang, A., DiMarco, E., Wilson, T., Siddiqui, M. S., Haq, I. ul, … Kishida, K. T. (2022). Intracranial subsecond dopamine measurements during a “sure bet or gamble” decision-making task in patients with alcohol use disorder suggest diminished dopaminergic signals about relief. *Neurosurgical Focus*. https://doi.org/10.3171/2022.11.FOCUS22614.

Liu, C., Tose, A. J., Verharen, J. P. H., Zhu, Y., Tang, L. W., de Jong, J. W., … Lammel, S. (2022). An inhibitory brainstem input to dopamine neurons encodes nicotine aversion. *Neuron*, *110*(18), 3018-3035.e7. https://doi.org/10.1016/j.neuron.2022.07.003

Loewenstein, G., Rick, S., & Cohen, J. D. (2008). Neuroeconomics. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.59.103006.093710

Ludvig, E. A., Sutton, R. S., & Kehoe, E. J. (2012). Evaluating the TD model of classical conditioning. *Learning and Behavior*. https://doi.org/10.3758/s13420-012-0082-6

Maes, E. J. P., Sharpe, M. J., Usypchuk, A. A., Lozzi, M., Chang, C. Y., Gardner, M. P. H., … Iordanova, M. D. (2020). Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. *Nature Neuroscience*, *23*(2), 176–178. https://doi.org/10.1038/s41593-019-0574-1

Markowitz, J. E., Gillis, W. F., Jay, M., Wood, J., Harris, R. W., Cieszkowski, R., … Datta, S. R. (2023). Spontaneous behaviour is structured by reinforcement without explicit reward. *Nature*, *614*(7946), 108–117. https://doi.org/10.1038/s41586-022-05611-2

Matsuda, W., Furuta, T., Nakamura, K. C., Hioki, H., Fujiyama, F., Arai, R., & Kaneko, T. (2009). Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.4029-08.2009

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey

positive and negative motivational signals. *Nature*. https://doi.org/10.1038/nature08028

McNamara, P., Durso, R., Brown, A., & Lynch, A. (2003). Counterfactual cognitive deficit in persons with Parkinson's disease. *Journal of Neurology Neurosurgery and Psychiatry*. https://doi.org/10.1136/jnnp.74.8.1065

Menegas, W., Babayan, B. M., Uchida, N., & Watabe-Uchida, M. (2017). Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. *ELife*. https://doi.org/10.7554/eLife.21886

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

Mohebi, A., Pettibone, J. R., Hamid, A. A., Wong, J. M. T., Vinson, L. T., Patriarchi, T., … Berke, J. D. (2019). Dissociable dopamine dynamics for learning and motivation. *Nature*, *570*(7759), 65–70. https://doi.org/10.1038/s41586-019-1235-y

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.16-05-01936.1996

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2008.03.006

Ostlund, S. B., & Maidment, N. T. (2012). Dopamine receptor blockade attenuates the general incentive motivational effects of noncontingently delivered rewards and reward-paired cues without affecting their ability to bias action selection. *Neuropsychopharmacology*. https://doi.org/10.1038/npp.2011.217

Ostlund, S. B., Wassum, K. M., Murphy, N. P., Balleine, B. W., & Maidment, N. T. (2011).

Extracellular dopamine levels in striatal subregions track shifts in motivation and response cost during instrumental conditioning. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.4759-10.2011

Pan, W. X., Coddington, L. T., & Dudman, J. T. (2021). Dissociable contributions of phasic dopamine activity to reward and prediction. *Cell Reports*, *36*(10). https://doi.org/10.1016/j.celrep.2021.109684

Papale, A. E., Stott, J. J., Powell, N. J., Regier, P. S., & Redish, A. D. (2012). Interactions between deliberation and delay-discounting in rats. *Cognitive, Affective and Behavioral Neuroscience*, *12*(3), 513–526. https://doi.org/10.3758/s13415-012-0097-7

Papale, A. E., Zielinski, M. C., Frank, L. M., Jadhav, S. P., & Redish, A. D. (2016). Interplay between Hippocampal Sharp-Wave-Ripple Events and Vicarious Trial and Error Behaviors in Decision Making. *Neuron*, *92*(5), 975–982. https://doi.org/10.1016/j.neuron.2016.10.028

Patriarchi, T., Cho, J. R., Merten, K., Howe, M. W., Marley, A., Xiong, W. H., … Tian, L. (2018). Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science*, *360*(6396). https://doi.org/10.1126/science.aat4422

Paxinos, G., & Franklin, K. B. J. (2001). The Mouse Brain in Stereotaxic Coordinates, 2nd edition. *Academic Press*, p. 360 p.

Poulin, J. F., Zou, J., Drouin-Ouellet, J., Kim, K. Y. A., Cicchetti, F., & Awatramani, R. B. (2014). Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. *Cell Reports*. https://doi.org/10.1016/j.celrep.2014.10.008

Rangel, A., & Clithero, J. A. (2013). The Computation of Stimulus Values in Simple Choice. In *Neuroeconomics: Decision Making and the Brain: Second Edition*. https://doi.org/10.1016/B978-0-12-416008-8.00008-5

Redish, A. D. (2016). Vicarious trial and error. *Nature Reviews Neuroscience 2016 17:3*, *17*(3), 147–159. https://doi.org/10.1038/nrn.2015.30

Redish, A. D., Sweis, B. M., Abram, S., Duin, A., Kazinka, R., Kocharian, A., … Thomas, M. (2021). Sunk cost sensitivity in mice, rats, and humans on the Restaurant Row and WebSurf tasks cannot be explained by attrition biases alone. *BioRxiv*, 2021.10.07.462802. https://doi.org/10.1101/2021.10.07.462802

Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. *Clasical Conditioning II: Current Research and Theory*.

Roese, N. J., & Olson, J. M. (1993). Self-Esteem and Counterfactual Thinking. *Journal of Personality and Social Psychology*, *65*(1), 199–206. https://doi.org/10.1037/0022-3514.65.1.199

Roese, N. J., Sanna, L. J., & Galinsky, A. D. (2012). The Mechanics of Imagination: Automaticity and Control in Counterfactual Thinking. In *The New Unconscious*. https://doi.org/10.1093/acprof:oso/9780195307696.003.0007

Salvador, A., Worbe, Y., Delorme, C., Coricelli, G., Gaillard, R., Robbins, T. W., … Palminteri, S. (2017). Specific effect of a dopamine partial agonist on counterfactual learning: Evidence from Gilles de la Tourette syndrome. *Scientific Reports*. https://doi.org/10.1038/s41598-017-06547-8

Sanna, L. J., Meier, S., & Turley-Ames, K. J. (1998). Mood, self-esteem, and counterfactuals: Externally attributed moods limit self-enhancement strategies. *Social Cognition*, *16*(2), 267–286. https://doi.org/10.1521/soco.1998.16.2.267

Saunders, B. T., Richard, J. M., Margolis, E. B., & Janak, P. H. (2018). Dopamine neurons create

Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nature Neuroscience*, *21*(8), 1072–1083. https://doi.org/10.1038/s41593-018-0191-4

Saunders, B. T., & Robinson, T. E. (2011). Individual variation in the motivational properties of cocaine. *Neuropsychopharmacology*, *36*(8), 1668–1676. https://doi.org/10.1038/npp.2011.48

Schmack, K., Bosc, M., Ott, T., Sturgill, J. F., & Kepecs, A. (2021). Striatal dopamine mediates hallucination-like perception in mice. *Science*, *372*(6537). https://doi.org/10.1126/science.abf4740

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*. https://doi.org/10.1016/j.neuron.2016.08.019

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience*, *13*(3), 900–913. https://doi.org/10.1523/jneurosci.13-03-00900.1993

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Shansky, R. M., & Murphy, A. Z. (2021, April 1). Considering sex as a biological variable will require a global shift in science culture. *Nature Neuroscience*, Vol. 24, pp. 457–464. https://doi.org/10.1038/s41593-021-00806-8

Sharpe, M. J., Chang, C. Y., Liu, M. A., Batchelor, H. M., Mueller, L. E., Jones, J. L., … Schoenbaum, G. (2017). Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nature Neuroscience*, *20*(5), 735–742.

https://doi.org/10.1038/nn.4538

Sheldon, K. M., Ryan, R., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/01461672962212007

Sokolowski, J. D., Conlan, A. N., & Salamone, J. D. (1998). A microdialysis study of nucleus accumbens core and shell dopamine during operant responding in the rat. *Neuroscience*. https://doi.org/10.1016/S0306-4522(98)00066-9

Solca, F., Poletti, B., Zago, S., Crespi, C., Sassone, F., Lafronza, A., … Ciammola, A. (2015). Counterfactual thinking deficit in Huntington's disease. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0126773

Sorg, B. A., & Kalivas, P. W. (1991). Effects of cocaine and footshock stress on extracellular dopamine levels in the ventral striatum. *Brain Research*. https://doi.org/10.1016/0006-8993(91)90283-2

Starkweather, C. K., Babayan, B. M., Uchida, N., & Gershman, S. J. (2017). Dopamine reward prediction errors reflect hidden-state inference across time. *Nature Neuroscience*. https://doi.org/10.1038/nn.4520

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*(7), 966–973. https://doi.org/10.1038/nn.3413

Steiner, A. P., & Redish, A. D. (2012). The road not taken: Neural correlates of decision making in orbitofrontal cortex. *Frontiers in Neuroscience*, (SEP). https://doi.org/10.3389/fnins.2012.00131

Steiner, A. P., & Redish, A. D. (2014). Behavioral and neurophysiological correlates of regret in

rat decision-making on a neuroeconomic task. *Nature Neuroscience*, *17*(7), 995–1002. https://doi.org/10.1038/nn.3740

Stelly, C. E., Haug, G. C., Fonzi, K. M., Garcia, M. A., Tritley, S. C., Magnon, A. P., … Wanat, M. J. (2019). Pattern of dopamine signaling during aversive events predicts active avoidance learning. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(27), 13641–13650. https://doi.org/10.1073/pnas.1904249116

Stopper, C. M., Khayambashi, S., & Floresco, S. B. (2013). Receptor-specific modulation of risk-based decision making by nucleus accumbens dopamine. *Neuropsychopharmacology*. https://doi.org/10.1038/npp.2012.240

Stott, J. J., & David Redish, A. (2014). A functional difference in information processing between orbitofrontal cortex and ventral striatum during decision-making behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2013.0472

Stuber, G. D., Hnasko, T. S., Britt, J. P., Edwards, R. H., & Bonci, A. (2010). Dopaminergic terminals in the nucleus accumbens but not the dorsal striatum corelease glutamate. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1754-10.2010

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*. https://doi.org/10.1007/bf00115009

Suzuki, S., Yamashita, Y., & Katahira, K. (2021). Psychiatric symptoms influence reward-seeking and loss-avoidance decision-making through common and distinct computational processes. *Psychiatry and Clinical Neurosciences*. https://doi.org/10.1111/pcn.13279

Sweis, B. M., Abram, S. V., Schmidt, B. J., Seeland, K. D., MacDonald, A. W., Thomas, M. J., & Redish, A. D. (2018). Sensitivity to “sunk costs” in mice, rats, and humans. *Science*,

361(6398), 178–181. https://doi.org/10.1126/science.aar8644

Sweis, B. M., Thomas, M. J., & Redish, A. D. (2018). Mice learn to avoid regret. *PLoS Biology*, *16*(6). https://doi.org/10.1371/journal.pbio.2005853

Takahashi, Y. K., Batchelor, H. M., Liu, B., Khanna, A., Morales, M., & Schoenbaum, G. (2017). Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards. *Neuron*, *95*(6), 1395-1405.e3. https://doi.org/10.1016/j.neuron.2017.08.025

Tiklová, K., Björklund, Å. K., Lahti, L., Fiorenzano, A., Nolbrant, S., Gillberg, L., … Perlmann, T. (2019). Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nature Communications*. https://doi.org/10.1038/s41467-019-08453-1

Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of Predicted Reward Omission by Dopamine Neurons in a Conditioned Inhibition Paradigm. *Journal of Neuroscience*, *23*(32), 10402–10410. https://doi.org/10.1523/jneurosci.23-32-10402.2003

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. *Science*. https://doi.org/10.1126/science.1105370

Tsai, H. C., Zhang, F., Adamantidis, A., Stuber, G. D., Bond, A., De Lecea, L., & Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*. https://doi.org/10.1126/science.1168878

Tye, K. M., Mirzabekov, J. J., Warden, M. R., Ferenczi, E. A., Tsai, H. C., Finkelstein, J., … Deisseroth, K. (2013). Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature*, *493*(7433), 537–541. https://doi.org/10.1038/nature11740

Ungless, M. A., Magill, P. J., & Bolam, J. P. (2004). Uniform Inhibition of Dopamine Neurons in the Ventral Tegmental Area by Aversive Stimuli. *Science*. https://doi.org/10.1126/science.1093360

Van Hoeck, N., Watson, P. D., & Barbey, A. K. (2015). Cognitive neuroscience of human counterfactual reasoning. *Frontiers in Human Neuroscience*, *9*(JULY), 1–18. https://doi.org/10.3389/FNHUM.2015.00420

Vandaele, Y., Lenoir, M., Vouillac-Mendoza, C., Guillem, K., & Ahmed, S. H. (2021). Probing the decision-making mechanisms underlying choice between drug and nondrug rewards in rats. *ELife*, *10*. https://doi.org/10.7554/ELIFE.64993

Wassum, K. M., Ostlund, S. B., Loewinger, G. C., & Maidment, N. T. (2013a). Phasic mesolimbic dopamine release tracks reward seeking during expression of pavlovian-to-instrumental transfer. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2012.12.005

Wassum, K. M., Ostlund, S. B., Loewinger, G. C., & Maidment, N. T. (2013b). Phasic mesolimbic dopamine release tracks reward seeking during expression of pavlovian-to-instrumental transfer. *Biological Psychiatry*, *73*(8), 747–755. https://doi.org/10.1016/j.biopsych.2012.12.005

Westbrook, A., van den Bosch, R., Määttä, J. I., Hofmans, L., Papadopetraki, D., Cools, R., & Frank, M. J. (2020). Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science*, *367*(6484), 1362–1366. https://doi.org/10.1126/science.aaz5891

Wickens, J. R., Horvitz, J. C., Costa, R. M., & Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1671-07.2007

Willmore, L., Cameron, C., Yang, J., Witten, I. B., & Falkner, A. L. (2022). Behavioural and dopaminergic signatures of resilience. *Nature*, *611*(7934), 124–132. https://doi.org/10.1038/s41586-022-05328-2

Witten, I. B., Steinberg, E. E., Lee, S. Y., Davidson, T. J., Zalocusky, K. A., Brodsky, M., … Deisseroth, K. (2011). Recombinase-driver rat lines: Tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron*. https://doi.org/10.1016/j.neuron.2011.10.028

Wyvell, C. L., & Berridge, K. C. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: Enhancement of reward "wanting" without enhanced "liking" or response reinforcement. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.20-21-08122.2000

Yip, S. W., Sacco, K. A., George, T. P., & Potenza, M. N. (2009). Risk/reward decision-making in schizophrenia: A preliminary examination of the influence of tobacco smoking and relationship to Wisconsin Card Sorting Task performance. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2009.01.012

Young, A. M. J., Joseph, M. H., & Gray, J. A. (1993). Latent inhibition of conditioned dopamine release in rat nucleus accumbens. *Neuroscience*. https://doi.org/10.1016/0306-4522(93)90378-S

Yun, M., Nejime, M., Kawai, T., Kunimatsu, J., Yamada, H., Kim, H. G. R., & Matsumoto, M. (2023). Distinct roles of the orbitofrontal cortex, ventral striatum, and dopamine neurons in counterfactual thinking of decision outcomes. *Science Advances*, *9*(32). https://doi.org/10.1126/SCIADV.ADH2831/SUPPL_FILE/SCIADV.ADH2831_SM.PDF

Zampese, E., Galtieri, D. J., Schumacker, P. T., & Surmeier, D. J. (2016). Determinants of

Selective Vulnerability of Dopamine Neurons in Parkinson's Disease. In *Handbook of Behavioral Neuroscience*. https://doi.org/10.1016/B978-0-12-802206-1.00041-6

Zeelenberg, M., & Pieters, R. (1999). Comparing service delivery to what might have Been: Behavioral responses to regret and disappointment. *Journal of Service Research*, *2*(1), 86–97. https://doi.org/10.1177/109467059921007

Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2018). Counterfactual Reasoning Underlies the Learning of Priors in Decision Making. *Neuron*, *99*(5), 1083-1097.e6. https://doi.org/10.1016/J.NEURON.2018.07.035

Appendix

Data Table 1. Comprehensive reporting of all main effects and interactions from ANOVA models, including those involving sex.

Figure	Sample	Comparison	Statistic	p value
1l: OZDM skipping probability (difference 30sec – 1 sec offers)	n=9 OZDM (6M/3F)	Interaction Rank Sex	F (3, 21) < 1 F (1.698, 11.89) = 5.060 F (1, 7) < 1	0.0297
1m: OZDM quitting probability (difference 30sec – 1 sec offers)	n=9 OZDM (6M/3F)	Interaction Rank Sex	F (3, 21) < 1 F (2.140, 14.98) = 19.93 F (1, 7) < 1	<0.0001
1n: WZDM skipping probability (difference 30sec – 1sec offers)	n=12 OZDM (6M/6F)	Interaction Rank Sex	F (3, 30) < 1 F (2.793, 27.93) < 1 F (1, 10) < 1	
1o: WZDM quitting probability (difference 30sec – 1 sec offers)	n=12 OZDM (6M/6F)	Interaction Rank Sex	F (3, 30) < 1 F (1.946, 19.46) = 1.093 F (1, 10) = 4.150	0.3532 0.0690
3c: mean reaction time on accepted trials	n=9 OZDM (6M/3F) n=12 WZDM 6M/6F)	Interaction DM Sex	F (1, 17) < 1 F (1, 17) = 8.720 F (1, 17) < 1	0.0089
4h: OZDM difference in mean dopamine to binned offers	n=9 OZDM (6M/3F)	Interaction Offer Length Sex	F (2, 14) < 1 F (1.006, 7.040) = 5.911 F (1, 7) < 1	0.0450
4j: WZDM difference in mean dopamine to binned offers	n=12 WZDM (6M/6F)	Interaction Offer Length Sex	F (2, 20) < 1 F (1.032, 10.32) = 5.271 F (1, 10) = 3.448	0.0430 0.0930
4m: Difference in mean dopamine to OZ outcomes	n=9 OZDM (6M/3F) n=12 WZDM (6M/6F)	OZ Outcome Decision Maker (DM) Sex OZ Outcome X DM OZ Outcome X Sex DM X Sex	F (1, 17) = 16.02 F (1, 17) < 1 F (1, 17) < 1 F (1, 17) = 1.184 F (1, 17) < 1 F (1, 17) < 1	0.0009 0.2918

		OZ Outcome X Sex X DM	F (1, 17) < 1	
4r: Difference in mean dopamine to WZ outcomes	n=9 OZDM (6M/3F) n=12 WZDM (6M/6F)	WZ Outcome Decision Maker (DM) Sex WZ Outcome X DM WZ Outcome X Sex DM X Sex WZ Outcome X Sex X DM	F (1, 17) = 24.30 F (1, 17) = 6.843 F (1, 17) = 1.043 F (1, 17) = 6.712 F (1, 17) < 1 F (1, 17) = 3.828 F (1, 17) = 3.232	0.0001 0.0181 0.3213 0.0190 0.0670 0.0900
6b: mean dopamine at earns	n=9 OZDM (6M/3F) n=12 WZDM 6M/6F)	Interaction DM Sex	F (1, 17) < 1 F (1, 17) < 1 F (1, 17) < 1	
7e: Mean change in dopamine before quit, binned by future value	n=9 OZDM (6M/3F)	Interaction Future Value Sex	F (1, 7) < 1 F (1, 7) = 6.284 F (1, 7) = 1.572	 0.0406 0.2502
7g: Mean change in dopamine before quit, binned by future value	n=12 WZDM (6M/6F)	Interaction Future Value Sex	F (1, 10) < 1 F (1, 10) = 4.815 F (1, 10) < 1	 0.0530
7i: Mean change in dopamine before quit, binned by future time	n=9 OZDM (6M/3F)	Interaction Future Value Sex	F (1, 7) < 1 F (1, 7) = 3.419 F (1, 7) = 1.038	 0.1069 0.3423
7k: Mean change in dopamine before quit, binned by future time	n=12 WZDM (6M/6F)	Interaction Future Value Sex	F (1, 10) < 1 F (1, 10) < 1 F (1, 10) < 1	
8c: area under the curve for high value and low value countdowns	n=9 OZDM (6M/3F) n=12 WZDM (6M/6F)	Value DM Sex Value x DM Value x Sex DM x Sex Value x DM x Sex	F (1, 17) = 9.394 F (1, 17) = 5.380 F (1, 17) < 1 F (1, 17) < 1 F (1, 17) < 1 F (1, 17) < 1 F (1, 17) = 2.932	0.0070 0.0331 0.1050
9c: mean change in dopamine post quit binned by past value	n=9 OZDM (6M/3F) n=12 WZDM (6M/6F)	Past Value DM Sex Past Value x DM	F (1.167, 19.84) = 12.69 F (1, 17) < 1 F (1, 17) = 1.092 F (2, 34) = 1.015	0.0013 0.3106 0.3733

		Past Value x Sex DM x Sex Past Value x DM x Sex	F (2, 34) < 1 F (1, 17) < 1 F (2, 34) = 1.935	0.1600
9d: slope of dopamine post quit binned by past value	n=9 OZDM (6M/3F) n=12 WZDM (6M/6F)	Past Value DM Sex Past Value x DM Past Value x Sex DM x Sex Past Value x DM x Sex	F (1.162, 19.75) = 15.43 F (1, 17) = 2.007 F (1, 17) < 1 F (2, 34) = 1.640 F (2, 34) < 1 F (1, 17) < 1 F (2, 34) < 1	0.0005 0.1746 0.2089
9g: mean change in dopamine post quit binned by past time	n=9 OZDM (6M/3F) n=12 WZDM (6M/6F)	Past Time DM Sex Past Time x DM Past Time x Sex DM x Sex Past Time x DM x Sex	F (1.310, 22.26) < 1 F (1, 17) < 1 F (1, 17) < 1 F (2, 34) <1 F (2, 34) < 1 F (1, 17) = 1.147 F (2, 34) = 1.555	 0.2991 0.2257
9h: slope of dopamine post quit binned by past time	n=9 OZDM (6M/3F) n=12 WZDM (6M/6F)	Past Time DM Sex Past Time x DM Past Time x Sex DM x Sex Past Time x DM x Sex	F (1.540, 26.19) = 1.111 F (1, 17) <1 F (1, 17) <1 F (2, 34) = 4.542 F (2, 34) = 1.706 F (1, 17) <1 F (2, 34) = 1.807	0.3299 0.0179 0.1967 0.1796
10l-m: Normalized mean DA for OZDM accepts and skips, binned by VTE	n=9 OZDM (6M/3F)	VTE OZ Outcome Sex VTE x OZ Outcome VTE x Sex OZ Outcome x Sex VTE x OZ Outcome x Sex	F (1, 7) = 3.406 F (1, 7) = 6.312 F (1, 7) < 1 F (1, 7) = 5.627 F (1, 7) < 1 F (1, 7) < 1 F (1, 7) < 1	0.1075 0.0403 0.0495
10n-o: Normalized mean DA for WZDM accepts and skips, binned by VTE	n=12 WZDM (6M/6F)	VTE OZ Outcome Sex VTE x OZ Outcome VTE x Sex OZ Outcome x Sex VTE x OZ Outcome x Sex	F (1, 10) < 1 F (1, 10) = 6.819 F (1, 10) = 2.428 F (1, 10) = 1.046 F (1, 10) = 2.448 F (1, 10) = 1.076 F (1, 10) < 1	 0.0260 0.1502 0.3304 0.1487 0.3241

11a: sum of path curvature during offer zone passes	n=9 OZDM (6M/3F) n=12 WZDM 6M/6F)	Interaction DM Sex	F (1, 17) < 1 F (1, 17) = 21.96 F (1, 17) < 1	0.0002
13e: zIdPhi with stimulation at less preferred flavors	n=7 OZDM (4M/3F) n=9 WZDM (6M/3F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 12) = 1.291 F (1, 12) < 1 F (1, 12) < 1 F (1, 12) = 10.29 F (1, 12) = 5.960 F (1, 12) < 1 F (1, 12) = 2.989	0.2781 0.0075 0.0311 0.1095
13f: zIdPhi with stimulation at more preferred flavors	n=7 OZDM (4M/3F) n=9 WZDM (6M/3F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 12) = 2.993 F (1, 12) = 2.156 F (1, 12) = 2.720 F (1, 12) < 1 F (1, 12) < 1 F (1, 12) < 1 F (1, 12) < 1	0.1092 0.1678 0.1250
13g: zIdPhi with stimulation at all flavors	n=7 OZDM (4M/3F) n=9 WZDM (6M/3F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 12) = 1.257 F (1, 12) < 1 F (1, 12) = 2.009 F (1, 12) < 1 F (1, 12) = 1.776 F (1, 12) < 1 F (1, 12) < 1	0.2841 0.1818 0.2074
13h: difference score of zIdPhi with stimulation at less preferred flavors	n=9 OZDM (6M/3F) n=12 WZDM 6M/6F)	Interaction DM Sex	F (1, 12) = 2.989 F (1, 12) = 10.29 F (1, 12) = 5.960	0.1095 0.0075 0.0311
13k: quit rate with stimulation at less preferred flavors	n=7 OZDM (4M/3F) n=9 WZDM (6M/3F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 12) < 1 F (1, 12) = 4.637 F (1, 12) < 1 F (1, 12) = 9.500 F (1, 12) = 1.043 F (1, 12) < 1 F (1, 12) = 1.991	0.0523 0.0095 0.3272 0.1836

13l: quit rate with stimulation at more preferred flavors	n=7 OZDM (4M/3F) n=9 WZDM (6M/3F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 12) < 1 F (1, 12) < 1 F (1, 12) = 1.243 F (1, 12) = 2.012 F (1, 12) < 1 F (1, 12) = 1.072 F (1, 12) < 1	 0.2867 0.1815 0.3209
13m: quit rate with stimulation at all flavors	n=7 OZDM (4M/3F) n=9 WZDM (6M/3F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 12) < 1 F (1, 12) = 2.194 F (1, 12) < 1 F (1, 12) < 1 F (1, 12) < 1 F (1, 12) < 1 F (1, 12) = 1.269	 0.1644 0.2820
13n: difference score of quit rate with stimulation at less preferred flavors	n=9 OZDM (6M/3F) n=12 WZDM 6M/6F)	Interaction DM Sex	F (1, 12) = 1.991 F (1, 12) = 9.500 F (1, 12) = 1.043	0.1836 0.0095 0.3272
14a: reaction time with stimulation at less preferred flavors	n=7 OZDM (4M/3F) n=9 WZDM (6M/3F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 12) = 1.483 F (1, 12) = 6.370 F (1, 12) < 1 F (1, 12) = 6.145 F (1, 12) = 6.711 F (1, 12) < 1 F (1, 12) = 2.107	0.2468 0.0267 0.0290 0.0236 0.1722
16d: mean signal with or without inhibition	n= 17 (12F/5M)	Interaction Stim Condition Sex	F (1, 15) < 1 F (1, 15) = 8.242 F (1, 15) < 1	 0.0117
17b: zIdPhi with inhibition at less preferred flavors	n= 7 OZDM (2M/5F) n=8 WZDM (1M/7F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 11) = 2.576 F (1, 11) = 1.688 F (1, 11) = 3.585 F (1, 11) = 1.566 F (1, 11) < 1 F (1, 11) < 1 F (1, 11) = 6.970e-006	0.1368 0.2205 0.0849 0.2368 0.9979
17c: zIdPhi with inhibition at more preferred flavors	n= 7 OZDM (2M/5F) n=8 WZDM	Stim Condition DM Sex	F (1, 11) < 1 F (1, 11) = 4.069 F (1, 11) = 3.355	 0.0687 0.0942

	(1M/7F)	Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 11) < 1 F (1, 11) < 1 F (1, 11) = 1.159 F (1, 11) < 1	0.3046
17d: zIdPhi with inhibition at all flavors	n= 7 OZDM (2M/5F) n=8 WZDM (1M/7F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 11) = 4.160 F (1, 11) = 1.120 F (1, 11) < 1 F (1, 11) = 4.582 F (1, 11) = 1.406 F (1, 11) < 1 F (1, 11) < 1	0.0662 0.3127 0.0555 0.2607
18b: quit rate with stimulation at more preferred flavors	n= 7 OZDM (2M/5F) n=8 WZDM (1M/7F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 11) < 1 F (1, 11) = 2.761 F (1, 11) < 1 F (1, 11) = 3.244 F (1, 11) = 3.262 F (1, 11) < 1 F (1, 11) = 3.140	0.1248 0.0991 0.0983 0.1040
18c: quit rate with stimulation at all flavors	n= 7 OZDM (2M/5F) n=8 WZDM (1M/7F)	Stim Condition DM Sex Stim Condition x DM Stim Condition x Sex DM x Sex Stim Condition x DM x Sex	F (1, 11) = 7.041 F (1, 11) = 2.446 F (1, 11) < 1 F (1, 11) < 1 F (1, 11) = 3.851 F (1, 11) < 1 F (1, 11) < 1	0.0224 0.1461 0.0755