2019

NEUROBEHAVIORAL MEASUREMENTS OF NATURAL AND OPIOID REWARD VALUE

Aaron Paul Smith
University of Kentucky, asm233@g.uky.edu
Author ORCID Identifier:
https://orcid.org/0000-0002-5417-014X
Digital Object Identifier: https://doi.org/10.13023/etd.2019.316

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Aaron Paul Smith, Student

Dr. Joshua S. Beckmann, Major Professor

Dr. Mark T. Fillmore, Director of Graduate Studies

NEUROBEHAVIORAL MEASUREMENTS OF NATURAL AND OPIOID REWARD VALUE

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the College of Arts and Sciences
at the University of Kentucky

By

Aaron Paul Smith

Lexington, Kentucky

Director: Dr. Joshua S. Beckmann, Associate Professor of Psychology

Lexington, Kentucky

2019

https://orcid.org/0000-0002-5417-014X

ABSTRACT OF DISSERTATION

NEUROBEHAVIORAL MEASUREMENTS OF NATURAL AND OPIOID REWARD VALUE

In the last decade, (non)prescription opioid abuse, opioid use disorder (OUD) diagnoses, and opioid-related overdoses have risen and represent a significant public health concern. One method of understanding OUD is as a disorder of choice that requires choosing opioid rewards at the expense of other nondrug rewards. The characterization of OUD as a disorder of choice is important as it implicates decision-making processes as therapeutic targets, such as the valuation of opioid rewards. However, reward-value measurement and interpretation are traditionally different in substance abuse research compared to related fields such as economics, animal behavior, and neuroeconomics and may be less effective for understanding how opioid rewards are valued. The present research therefore used choice procedures in line with behavioral/neuroeconomic studies to determine if drug-associated decision making could be predicted from economic choice theories. In Experiment 1, rats completed an isomorphic food-food probabilistic choice task with dynamic, unpredictable changes in reward probability that required constant updating of reward values. After initial training, the reward magnitude of one choice subsequently increased from one to two to three pellets. Additionally, rats were split between the Signaled and Unsignaled groups to understand how cues modulate reward value. After each choice, the Unsignaled group received distinct choice-dependent cues that were uninformative of the choice outcome. The Signaled group also received uninformative cues on one option, but the alternative choice produced reward-predictive cues that informed the trial outcome as a win or loss. Choice data were analyzed at a molar level using matching equations and molecular level using reinforcement learning (RL) models to determine how probability, reward magnitude, and reward-associated cues affected choice. Experiment 2 used an allomorphic drug versus food procedure where the food reward for one option was replaced by a self-administered remifentanil (REMI) infusion at doses of 1, 3 and 10 μg/kg. Finally, Experiment 3 assessed the potential for both REMI and food reward value to be commonly scaled within the brain by examining changes in nucleus accumbens (NAc) Oxygen (O_2) dynamics. Results showed that increasing reward probability, magnitude, and the presence of reward-associated cues all independently increased the propensity of choosing the associated choice alternative, including REMI drug choices. Additionally, both molar matching and molecular RL models successfully parameterized rats' decision dynamics. O_2 dynamics were generally commensurate with the idea of a common value signal for REMI and food with changes in O_2 signaling scaling with the reward magnitude of REMI rewards. Finally, RL model-derived reward prediction errors significantly correlated with peak O_2 activity for reward delivery, suggesting a possible

neurological mechanism of value updating. Results are discussed in terms of their implications for current conceptualizations of substance use disorders including a potential need to change the discourse surrounding how substance use disorders are modeled experimentally. Overall, the present research provides evidence that a choice model of substance use disorders may be a viable alternative to the disease model and could facilitate future treatment options centered around economic principles.

KEYWORDS: Choice, Remifentanil, Matching, Reinforcement Learning, Reward Prediction Error, Nucleus Accumbens

Aaron Paul Smith

(Name of Student)

July 1st, 2019

Date

NEUROBEHAVIORAL MEASUREMENTS OF NATURAL AND OPIOID REWARD VALUE

By

Aaron Paul Smith

Joshua S. Beckmann, Ph.D.
Director of Dissertation

Mark T. Fillmore, Ph.D.
Director of Graduate Studies

July 1st, 2019
Date

ACKNOWLEDGMENTS

This dissertation would not have been possible without the assistance and support of several people. First, my graduate mentor, Dr. Joshua S. Beckmann. Our many conversations about the philosophy of science have truly given me an appreciation for what science is and could be. Furthermore, you have facilitated building my confidence to push the bounds of my own research even into areas I may initially know very little about. I would also like to thank my former graduate mentor, Dr. Thomas Zentall, who has helped develop my skills as a scientist through many meetings of theory crafting. I am also grateful to my dissertation committee members, Dr. Michael Bardo and Dr. Josh Lile, for providing insight and assistance at various points along the journey to complete the current dissertation and Dr. Joseph Benitez for serving as the outside examiner. The completion of this dissertation was also aided by the technical assistance of Josh Lavy as well as support from my friend and colleague, Dr. Andrew T. Marshall, who has helped me through many frustrating coding barriers since my undergraduate career. The present thesis would also not have been possible without the rats that were included in the present set of experiments.

I am also indebted to my friends and family for their support and encouragement throughout my graduate career. To my wife, Nicole Smith. The daily support, encouragement, and love you have given me cannot be understated. Thank you for taking this journey with me from a small Kansas town of 200 people to a town of 300,000 people 600 miles away from home. Finally, to my parents, Scott and Deb Smith. My journey in science may have kept me away (I'm sorry I don't call more), but your support has been a constant source of morale.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

VALUE AND ITS MEASUREMENT

Functional Consequences in Decision Sciences

A significant portion of the scientific literature is aimed at studying how to attenuate maladaptive behavior. The behavior under study ranges from topics such as getting people to exercise more, avoid unhealthy diets, stop using illicit substances, prepare for long-term financial needs, and not spend money beyond their current means. There is no doubt that many factors separate the propensity of one maladaptive behavior from another. However, a common unifying mechanism is the choice to engage in one action at the exclusion of others.

If choice is indeed a unifying mechanism of maladaptive behavior, then the decision-making processes leading to choice are implicated as potential therapeutic targets. For example, one of the most prominent paradigms in which decision making has found predictive validity is delay discounting (Odum, 2011). Delay discounting describes the decreased propensity for individuals to choose increasingly temporally distant rewards in favor of more proximal ones. Interestingly, the impulsive choice propensity characterized by delay discounting has shown remarkable clinical relevance through associations with behavior such as illicit substance abuse (Mackillop et al., 2011), substance abuse-related treatment outcomes (Loree, Lundahl, & Ledgerwood, 2015), gambling (Alessi & Petry, 2003), the incurrence of debt (Beauchaine, Ben-David, & Sela, 2017), attention-deficit hyperactivity disorder (Patros et al., 2016), and obesity (Rasmussen, Lawyer, & Reilly, 2010). However, perhaps more important was the functional characterization of delayed rewards becoming less valuable through time. Once known, interventions aimed at improving temporal perceptions (Smith, Marshall, & Kirkpatrick, 2015), delay tolerance (Renda & Madden, 2016), working memory (Renda, Stein, & Madden, 2015), and future orientation (Bulley & Gullo, 2017) became novel

therapeutic strategies to reduce impulsive choice and, ostensibly, other maladaptive choice behaviors.

Impulsive choice serves as a prominent example for how basic decision sciences can reveal potential therapies for clinically relevant disorders. However, just as choice can serve as a unifying process for many disorders, the hypothetical construct of reward value can serve as a mediating mechanism across different forms of choice.

Reward valuation is a prominent point of study for the emerging sub disciplines of behavioral economics and neuroeconomics seeking to understand clinically relevant conditions under the umbrella term of *reward pathology* (e.g., Jarmolowicz, Reed, DiGennaro Reed, & Bickel, 2016). Reward pathology generally refers to a characterization of choices thought to be mediated by a disproportional valuation of one choice alternative over others. For instance, in delay discounting, an impulsive chooser is thought to overvalue rewards offered immediately or, inversely, undervalue future rewards.

Brief History of the Concept of Value

Our current understanding about reward valuation stems from classical economic thought traditionally ascribed to renaissance economist Blaise Pascal (Glimcher, 2011; Glimcher, Camerer, Fehr, & Poldrack, 2013). Pascal was attempting to understand how people should make decisions under uncertainty. For example, should an individual pay 45 florins for a ticket that provides a 50% chance at 100 florins? To answer decisions involving uncertainty, Pascal proposed the idea of expected value (EV), or the expected amount a decision returns on average. Shown in Equation 1, the EV is calculated by multiplying the magnitude of a gain or loss (x_i) by the probability of its occurrence (p_i).

$$EV(x) = \sum_i x_i p_i(x_i) \quad (1)$$

Thus, purchasing the ticket in the above example is a wise decision, as it produces an expected value of 50 florins. Pascal's contribution is notable as it is the first to ascribe what decision makers *should* do, otherwise known as normative decision making.

Although EV was an important step towards understanding decision making, it became apparent to Daniel Bernoulli that it was an incomplete account (Bernoulli, 1738/1954; c.f. Glimcher, 2011). Bernoulli noted that not all individuals appeared to behave as though absolute increases in reward amount translated to equivalent increases in reward value. For instance, imagine a beggar comes upon a ticket that offers a 50% chance of winning 20,000 florins. A wealthy individual notices the beggar's ticket and then offers the beggar a certain 7,000 florins for the ticket. According to normative theory at the time, the beggar should refuse the ticket as it had an EV of 10,000 florins, or 3,000 more than the offer. However, if the beggar chooses to accept the certain 7,000 florins, then it is not necessarily irrational, as it would mean both meal and shelter security for some time to come.

To account for why the beggar might accept the offer, Bernoulli suggested the idea of *utility* loosely described as *subjective value*. To get at the idea of the hidden variable utility, Bernoulli computed the logarithm of an outcome's amount to derive what he referred to as *utils*. When Bernoulli's calculations are applied, the adjusted EV value of the beggar's ticket becomes 2.15 utils (i.e., 0.5 * log [20,000]) which is less than the wealthy man's offer of 3.8 utils (i.e., 1.0 * log [7,000]). Thus, a decision to take the certain 7,000 becomes justified. Stated differently, the utility of the beggar gaining a certain 7,000 was greater than the possibility of getting 3,000 more florins (on average); this is because rather than more florins producing a linear increase in reward value (as in Pascal's case), an increase in florins produces marginal gains reflective of a concave (or negatively accelerating) utility function. The translation of Bernoulli's subjective utility today (Glimcher et al., 2013) has been generalized to forms similar to Equation 2

$$EU(x) = \sum_i u_i(x_i)p_i(x_i) \tag{2}$$

in which the probability of a given outcome ($p_i(x_i)$) is multiplied by the outcome's scaled utility ($u_i(x_i)$) to produce expected utility (EU).

The notion of EU helped describe many violations of EV when the choices were isomorphic, or between commodities of the same type (cf. Glimcher, 2011; Glimcher et al., 2013). Once the appropriate utility function for a given set of choices was established, it could be used to predict future choices. However, an inherent problem with the idea of

the utility function was in the scaling of how valuable one commodity was relative to a different commodity, also known as allomorphic choice (Pareto, 1906/1971). Such scaling is difficult is because EU is not observable and only revealed through observed preferences. For instance, consider an example from Glimcher (2011) in which a participant prefers apples to oranges to grapes and we assign each fruit utilities of 3, 2, and 1, respectively. If our hypothetical person has two of each fruit, then they have a combined utility of 12, but we want to raise it to 16. One simple solution would be to add some combination of oranges and grapes to evenly raise the combined utility to 16. However, what if further preference tests reveal a new ordering of apples > oranges > pears > grapes > apricots. Suddenly, the assignment of utilities would necessarily change to perhaps 5, 4, 3, 2, and, 1. Thus, the potential combination of fruits to raise our participant's overall utility to 16 is now altered.

The issue revealed by the above example is that measures of reward value do not operate at the level of a cardinal scale. Stated differently, knowing that our participant prefers oranges to grapes is an ordinal relationship, but it is unknown if oranges (2 utils) truly offer twice the utility of grapes (1 util). By extension, knowing that oranges have a utility of two from a model does not allow for the prediction of preferences against a yet untested commodity. If a cardinal scale were existent, one could theoretically know oranges have a "True" value of 2 utils that would be invariant and could be compared against preferences for other commodities not assessed in the same context. The limitations of utility are also further evidenced by more recent research showing that preferences can be influenced by the dimension on which someone is asked to decide (e.g., wins vs. losses) and the framing of information; these effects on decision making and others are often referred to as the contextual dependency of decision-making (e.g., Birnbaum, 1992; Kahneman & Tversky, 1979). Thus, alluding back to the example of delay discounting, simply knowing participants preferred temporally proximal rewards over distant ones was not the primary outcome of the studies. Rather, it was the understanding that rewards became less valuable as they were delayed through time (that is, the functional consequence) that informed potential treatment strategies. In the same way, future studies seeking to understand reward value will need to uncover the

functional consequences that modulate reward value rather than solely compiling a list of preferences amongst commodities.

The Value of Drugs of Abuse

A prominent area of research that often invokes the idea of reward value is in substance abuse research. Illicit drug use is a pervasive public health concern estimated to cost the US $193 billion annually (National Drug Intelligence Center Product, 2011). Furthermore, the US is experiencing an opioid crisis stemming in part from the over-prescription of opioid analgesics. In two decades, the number of annually dispensed opioids nearly tripled to $248 million (Dart et al., 2015). Of those receiving opioids, 21-29% misuse them, 8-12% develop opioid use disorder (OUD; Vowles et al., 2015), and 4-6% transition to illicit opioid use (Carlson, Nahhas, Martins, & Daniulaityte, 2016). Additionally, 12.5 million people over the age of 12 reported past-year opioid misuse, 2.4 million people met OUD criteria (SAMHSA, 2017), and the mean opioid overdose deaths nearly tripled since 2002 to an average of 91 per day across the US (CDC, 2017). Thus, there is substantial public health cost of opioid use that highlights the need for treatment and prevention strategies.

Substance use disorders (SUDs) like OUD have also been described as “disorders of choice” (Kalivas, Volkow, & Seamans, 2005; Lamb, Maguire, Ginsburg, Pinkston, & France, 2016), in that individuals must choose the drug reward at the expense of other nondrug alternatives. Indeed, the notion of drug choices at the expense of nondrug alternatives is also reflected in many DSM-V psychiatric criteria for OUD (Association, 2013). Thus, decision-making processes, including reward-value mechanisms, have the potential to be therapeutic targets for reducing illicit drug use. Some laboratories have embraced the choice mechanism of decision making (e.g., Ahmed, 2010; Banks & Negus, 2012), but many use what is here referred to as single schedules of reinforcement (e.g., Belin-Rauscent, Fouyssac, Bonci, & Belin, 2016; Jones & Comer, 2013; Marchant, Li, & Shaham, 2013; N Kearns, A Gomez-Serrano, & J Tunstall, 2011; Sanchis-Segura & Spanagel, 2006), as described below

Common Measures of Reward Value in Substance Abuse Research

Conditioned place preference (Spragg, 1940) is perhaps the most ubiquitously used preclinical procedure (Childs & de Wit, 2009 in humans) to study the rewarding efficacy of a drug without having individuals actually self-administer (Bardo & Bevins, 2000). In a typical 3-chamber procedure consisting of two side chambers adjoined by a central chamber, an animal is initially placed in the central compartment and allowed access to the two side chambers. The time that the animal spends in the two separate rooms is measured as a baseline. Then, at separate time points, the animal is injected with either a dose of drug or saline and restricted to either the drug- or saline-paired room, which are made to be as distinct from one another as possible. Finally, after a number of conditioning sessions for both drug and saline, the animal is again placed in the central compartment and allowed free access to both rooms. The primary outcome measure is the time spent in each room. If the time spent in the drug-paired chamber is increased (both compared to the saline-paired chamber and baseline), then the results are interpreted as a binary distinction of the drug having reward value.

While conditioned place preference can assess drug reward value through noninvasive procedures, having an animal actually choose and take drug reward provides greater face validity as it reflects what human drug users do. As such, self-administration is generally referred to as the gold standard for assessing drug-associated reward value (Weeks, 1962). In animals, self-administration occurs via implantation of a catheter into the animal's vein that allows for direct intravenous infusions of drug. Self-administration procedures typically use a constant response requirement, otherwise known as a fixed-ratio (FR) schedule, on an operant such as a lever that, once completed, allows for self-administration of a drug. The amount (or rate) of intake is then interpreted as an index of how valuable or rewarding the drug is to a particular individual (Belin-Rauscent et al., 2016).

Further methodologies have also been developed on top of self-administration to striate individuals into more or less addiction-prone categories. One of the earliest was the progressive ratio schedule (Griffiths, Findley, Brady, Dolan-Gutcher, & Robinson, 1975; Hodos, 1961), in which the response requirement for each subsequent self-

administered dose is increased either trial-by-trial or across sessions (Richardson & Roberts, 1996). The primary measure with progressive ratio is the breakpoint, or the last completed FR followed by a predetermined period of time during which no responding occurs. Relatively higher breakpoints are interpreted as the individual exerting more effort to earn the reward and, by extension, evidence of the reward providing more utility. Another variant is the escalation model (Ahmed, Walker, & Koob, 2000), where rats are given either short (~1-3 hr) or long (~ 6 hr) access to self-administer a drug. Rats given extended 6-hr access will typically increase their rate of intake beyond the short access group and show resistance to drug extinction or continued use despite concomitant punishment (e.g., shock; Ahmed & Koob, 1998; Edwards & Koob, 2013). The escalation of drug intake and resistance to punishment ostensibly represents a divergence from reward-based decision making to compulsive decisions for drug reward, regardless of consequences.

A final paradigm is an adaptation of the economic demand model (Hursh & Silberberg, 2008). In demand procedures, the value of a commodity is assessed by determining an individual's response rate for a given commodity across different levels of an assumed unitary scale, *unit price*. Unit price is defined as the cost-benefit ratio of the reward magnitude divided by its cost (e.g., response requirement) that is manipulated across different experimental components. A demand equation, such as the exponential form shown in Equation 3, is then fit to the data.

$$Q = Q_0 + k(e^{-\alpha P} - 1) \quad (3)$$

In Equation 3, Q_0 is the value of a commodity when it is freely available (or at least at minimum price), α determines the rate of decline in response rates as a function of unit price (or a commodity's elasticity), and k is a scalar constant. While similar to progressive ratios, demand equations fit rates of intake as a function of unit price rather than only looking at the FR where responding stops. Commodities with relatively higher α parameters are said to be more elastic and maintain less reward value as a function of increases in price. Therefore, a commodity of greater reward value may have a relatively higher Q_0 and/or lower α. The demand equation can suitably parametrize several different drug commodities (Bentzley, Fender, & Aston-Jones, 2013; Hursh & Silberberg, 2008), and there is some evidence that testing two commodities in isolation can predict ordinal

preferences in choice (Johnson & Bickel, 2006; Kearns, Kim, Tunstall, & Silberberg, 2017; Schwartz, Kim, Silberberg, & Kearns, 2017; but see Schwartz, Silberberg, Casey, Paukner, & Suomi, 2016).

Theoretical and Pragmatic Issues with Single Schedule Measures

Rate Measures are Dissociable from Preferences

While there are merits to the above approaches, each procedure, when used to quantify the reward value of a single commodity (i.e., a single schedule), suffers from similar pragmatic and theoretical issues. One issue is the assumption that response rates serve as a valid measure for the utility of a commodity (Banks & Negus, 2012). Economic determination of reward value uses choice as the gold standard, and rates are often dissociated from choice preferences (e.g., Bonem & Crossman, 1988; Picker & Poling, 1982). Perhaps the most recognizable dissociation between rates and preference comes from the self-administration dose-response curve. In single schedules, increasing doses often produces a bitonic, "inverted U-shape" in response rates (e.g., Katz, 1989; Mello & Negus, 1996; Weeks, 1962), where intermediate dose values produce the highest rates of responding. Thus, one would conclude that the intermediate doses are the most rewarding due to the elevated rates of responding. The logic for higher doses decreasing rates is often attributed to aversive and nonspecific effects (Katz, 1989). However, when individuals are offered a choice between dose values and some other commodity (e.g., food) or a higher dose of the drug, the higher dose often produces larger drug preferences (Beckmann, Chow, & Hutsell, 2019; Freeman & Woolverton, 2011; Hutto & Crowder, 1997; Johanson & Schuster, 1975; Negus, 2006). Additionally, individuals described as 'compulsive' drug users that show escalated drug use in single schedules often do not show corroborative drug preference changes in choice procedures (Caprioli, Zeric, Thorndike, & Venniro, 2015; Lenoir, Serre, Cantin, & Ahmed, 2007; Schwartz et al., 2017; c.f. Lenoir, Cantin, Vanhille, Serre, & Ahmed, 2013; Negus & Rice, 2009). Such dissociation suggests that rate measures of reward efficacy are not in agreement with choice procedures and are susceptible to other factors independent of the value of the commodity here called *direct effects* (Katz, 1989).

Direct Effects

Two notable direct effects apparent from single schedules are satiation and motoric effects. Satiation refers to the hypothesis that the decrease in rates of responding at relatively higher doses is due to an individual needing fewer infusions to reach threshold levels of reward. Evidence for satiation comes from findings that animals will titrate their intake and show stable inter-reward-intervals as a function of dose (Katz, 1989), and that higher doses can lead to increased latencies to respond in choice procedures without concomitant decreases in drug preference (Beckmann et al., 2019). Additionally, forced time-out periods between doses of self-administered cocaine can change the peak of the dose-response curve. Short timeouts shift the peak of the function to lower doses (Caine & Koob, 1994), while longer timeouts do not show the decreased rates or inverted U-shaped curve typical at higher doses (Griffiths, Bradford, & Brady, 1979). Finally, relatively large amounts of food (Goldberg, 1973) and intracranial self-stimulation voltage (Reynolds, 1958) can also produce decreases in response rates similar to a dose-response curve. Thus, some of the apparent decreases in response rates may be due to a satiety effect of the individual spacing their reward intake that, in temporally constrained procedures such as single schedules, appears as a reduction in value.

The issue of motoric effects has two components: the rate-altering effects of some drugs and rate-dependent effects. In terms of gross motoric effects, stimulants (Antoniou, Kafetzopoulos, Papadopoulou-Daifoti, Hyphantis, & Marselos, 1998; Witkin, 1993) and opioids (Browne & Segal, 1980; Smith, Greene-Naples, Lyle, Iordanou, & Felder, 2009) can increase or decrease locomotor activity that is both temporally dependent and can change between acute versus chronic administration. Thus, when an individual's ability to move is directly related to the measure of value, compounds that can affect motoric ability confound motoric and valuation processes. The potentially confounding motor effects are also a primary concern whenever a treatment is applied (be it pharmacological, genetic, or anatomical), as it can be difficult to dissociate any observed rate changes as being due to altered reinforcing efficacy of the scheduled commodity from a motoric-altering effect of the treatment (Mello & Negus, 1996).

The difficulty of dissociating motoric from preference effects is also further compounded by rate dependency (Branch, 1984). Rate dependency describes an

interaction of rate-altering effects with an individual's baseline performance. Rate increases are more probable at lower baseline rates while decreases more probable with higher baselines. To demonstrate rate dependency, researchers have manipulated the baseline rate of responding through using either interval or ratio schedules (Kelleher & Morse, 1964; Lucki, 1983; McMillan, 1969; Thompson, Honor, Verchota, & Cleary, 1984) that tend to produce lower and higher rates of responding, respectively (Powell, Honey, & Symbaluk, 2016). For example, individuals experienced with both schedules that are challenged by the same doses of amphetamine show divergent increases and decreases in rates depending on whether the baseline rate of performance was lower or higher, respectively (e.g., Kelleher & Morse, 1964). Importantly, the effects of rate dependency are contingent upon the baseline rate of behavior and not necessarily the type of schedule used (McMillan, 1969). Thus, the issue of motoric effects confounding single schedule measures of value is robust and makes comparisons of both within- and between-subject manipulations difficult to interpret.

Inherent Invocation of the Cardinal Scale

A final more theoretical issue with single schedules is the lack of recognizing the contextual dependency of a commodity's value by offering it in isolation. The contextual-dependency of decisions has long been recognized in economics (e.g., Glimcher, 2011; Kahneman & Tversky, 1979) and represents an important theoretical assumption: choice preferences are relative to other available commodities. That single schedules offer reward commodities in isolation implies an assumption that the obtained measure of reward value is indicative of an inherent or "True" value, similar to the idea of the cardinal scale. However, from the earlier example of assigning utility values to different fruits, it is known that the idea of a cardinal scale is flawed.

One prominent invocation of cardinal scale ideology (despite recognizing the variability of reward value) is the use of the *essential value* metric stemming from economic demand. Essential value (Hursh & Silberberg, 2008) describes the elasticity of a commodity (α) across the assumed unidimensional space of unit price after normalizing for differences in maximum consumption (Q_0; Equation 3). The goal of the essential value metric is to create a single measure of reward value that can compare commodities across similar procedures (e.g., Christensen, Silberberg, Hursh, Huntsberry, & Riley,

2008; Schwartz et al., 2019; Smethells, Harris, Burroughs, Hursh, & LeSage, 2018) and be used to identify 'addiction-like' vulnerabilities in decision-making (e.g., Bentzley, Jhou, & Aston-Jones, 2014; Murphy, MacKillop, Skidmore, & Pederson, 2009).

There are multiple problems with the idea that reward value as a function of unit price (i.e., essential value) is representative of an absolute measurement of value. First, as already mentioned, essential value invokes the idea of an invariant value measure that can be placed upon a cardinal scale from which to compare all commodities against. Second, the dimension of unit price fails to take into account that different dimensions (e.g., effort, time, magnitude, probability, etc.) are not psychophysically scaled the same (Stevens, 1957). For example, unit price is defined as the magnitude of a reward divided by its cost (except cost is ubiquitously effort). If unit price indeed represents a unidimensional space, then manipulating either the cost or magnitude of a commodity, such that unit prices are equivalent, should result in equivalent measurements of elasticity. However, manipulating cost or magnitude produces differences in essential value estimates (Smith, Rupprecht, Sved, & Donny, 2016) and subjects' choices between commodities of identical unit prices (but varying in cost or magnitude) are not indifferent (Madden, Bickel, & Jacobs, 2000). Essential value is also altered by the presence of a concurrently available commodity (Carroll & Rodefer, 1993; Smethells et al., 2018) and the overall length of time that a commodity is available (Carroll & Rodefer, 1993; Foster, Kinloch, & Poling, 2011). Thus, unit price is not a singular dimension for comparing reward value, and the metric of essential value is dependent upon the decision-making context.

While the above inconsistencies are troubling for essential value, they are also predictable from a differentially scaled multidimensional view of reward value. Differential scaling suggests that a change in one dimension (e.g., effort) may have different psychophysical effects than the addition of one food reward (i.e., magnitude), despite the same unit price. Therefore, inconsistencies in manipulations of magnitude or effort that return different essential value measurements are not without theoretical cause. Furthermore, manipulations of different commodity-relevant dimensions can have independent effects on choice preferences (Arvanitogiannis & Shizgal, 2008; Beckmann et al., 2019; Smith, Beckmann, & Zentall, 2017a), which are not possible under the unit

price unidimensional assumption. Indeed, the inconsistencies in effects with the demand paradigm point to the necessity of a multidimensional view (described in detail below).

A distinction should be made between demand as a subdomain of economics and the essential value metric. The above section denotes issues with the use of essential value as a unifying metric for reward value; these same issues do not necessarily extend to broader applications of demand. Indeed, demand can serve as a useful behavioral economic tool (Bickel, Marsch, & Carroll, 2000) grounded by economic principles (Rachlin, Battalio, Kagel, & Green, 1981; Rachlin, Green, Kagel, & Battalio, 1976). Furthermore, demand can aid in understanding choice mechanisms such as cross-price elasticity assessments of substitutes and complements (Rachlin et al., 1976). However, researchers using essential value or single schedules as metrics of reward value should appreciate the shortcomings of single schedule measures, the contextual dependency of reward value, and the multidimensional nature of reward value less they become overconfident in their measures' predictive efficacy.

Multidimensional Scaling of Reward Value: The Progression of Matching

One reason why single--schedule paradigms might persist despite the noted issues is hysteresis. Rates of behavior were originally thought to be a measure of reward strength (e.g., Skinner, 1932; Skinner, 1938), and some have posited that differences between rate and choice measures might be reflective of multiple decision systems or independent aspects of reward value (e.g., Berridge, Robinson, & Aldridge, 2009; Regier & Redish, 2015). A discussion of the latter point would extend beyond the scope of this manuscript. Rather, this section is meant to describe an alternative approach from the experimental analysis of behavior (EAB) literature that was also steeped in the notion of response rates as reward strength. However, upon meeting theoretical challenges similar to those noted above, the EAB field altered their formulations of reward value as told through the progression of the matching equations.

Tautological Matching

Matching was originally derived from the observation that response rates came under orderly control of relative schedules of reinforcement (Davison & McCarthy,

1988). In a typical matching experiment, subjects can choose to allocate responses to two operants that reward responding at different (and locally variable) rates, also known as variable interval (VI) schedules. Under concurrent VI schedules, Herrnstein (Herrnstein, 1961) originally noted that the relative response rates of pigeons allocated their responding according to the relative reinforcement obtained on each key. Herrnstein quantified the effect using Equation 4

$$\frac{B_1}{B_2} = \frac{R_1}{R_2} \tag{4}$$

where *B* refers to the absolute behavior (response rates) between choice alternatives and *R* refers to the obtained reinforcers for choice 1 and 2. Although originally formulated for relative rates of responding and reinforcers, both sides of the matching equation were easily expanded to other dimensions. For example, subsequent work showed that relative responding matched dimensions such as relative delay and magnitude of reward while relative rates could be substituted for relative time allocation (Davison & McCarthy, 1988).

While Equation 4 does well to describe behavior under a variety of conditions, it soon needed alterations stemming from pragmatic issues in the use of rates as the output measure and in the extension to single schedules or operants (c.f., Davison & McCarthy, 1988; Herrnstein, 1970, 1974). Herrnstein originally thought that response rates were a linear function of reward rates. However, Catania (1963a, 1963b) showed non-linearity between response and reinforcer rates due in part to an organism having a maximum motoric output. Thus, any subsequent formulations typically took on a hyperbolic form to allow for an asymptote in responding. A second extension was also needed to account for the fact that responding on scheduled operants was not the only behavior available to an individual. Indeed, typically non-measured 'leisure behavior' was recognized in matching research and encompassed a class of any behavior other than responding for the scheduled commodity (B_e). Rather than simply being left out of the equations as error, leisure behavior was thought of as choice behavior (Herrnstein, 1970). The non-linear functionality between rates and reinforcers as well as leisure behavior were encompassed in Equation 5.

$$\frac{B_1}{B_e \ldots + B_N} = \frac{kR_1}{R_e \ldots + R_N} \quad (5)$$

Colloquially known as Herrnstein's hyperbola, Equation 5 suggests the rate of responding for a behavior (B_1, B_2... B_N) is now a function of its rate of reinforcement scaled according to an organism's maximal rate of responding for a given operant (k) divided by the rate of reinforcement for all other extraneous reinforcement (R_e) and other concurrently available choices (R_N). Note that while the use of response rate as a dependent variable here suffers from similar issues noted above, scaling behavior according to an individual's own motoric output takes one step towards trying to rectify the influence of extraneous variables other than experimenter manipulated variables (e.g., reward rate).

Generalized Matching

Although Herrnstein's hyperbola was a respectable step in attempting to control for confounding issues in rate measures of behavior, it was not consistent with empirical findings (Davison & McCarthy, 1988; McDowell, 2005, 2013). Namely, estimates of k and R_e were not always independent or invariant. Herrnstein's hyperbola was therefore replaced by the somewhat less theoretical generalized matching law (Baum, 1974) shown in Equation 6.

$$\frac{B_1}{B_2} = b\left(\frac{R_1}{R_2}\right)^s \quad (6)$$

Equation 6 was the beginning of a more modern class of matching equations called power functions (McDowell, 2005) that suggest that behavior (B) is a function of reinforcement rate (R) scaled with a power term (s) and multiplied by a bias term (b). The use of bias and sensitivity account for deviations from matching similar to Herrnstein's attempts, but in somewhat less of a theoretical manner using modeled free parameters. However, the sensitivity term (s) can be interpreted similar to power functions from previously-mentioned psychophysical studies (e.g., Stevens, 1957) as an individual's sensitivity to differences between and changes in relative rates of reinforcement. However, the behavior of interest needs to be measured under multiple levels of reinforcement rates to obtain theoretically accurate estimates of s. For tautological matching to occur, estimates of s and b should approximate values of one indicating that the relative response behavior

closely followed changes in relative reward rates and was without bias for either choice alternative. Despite some limitations of interpretation, the generalized matching equation has shown efficacy in describing a range of (non)human data from primarily concurrent VI choice studies (e.g., Baum & Rachlin, 1969; Wearden & Burgess, 1982) and can be an effective tool to describe how sensitive individuals' preferences are to changes in experimenter-manipulated variables.

The progression of matching was still not finished. As noted earlier, multiple dimensions can affect behavior (Baum & Rachlin, 1969; Rachlin, 1971). To account for multiple dimensions affecting behavior, concatenation of each was suggested as shown in Equation 7.

$$\frac{B_1}{B_2} = b * \frac{R_1}{R_2} * \frac{M_1}{M_2} * \frac{I_1}{I_2} * \frac{\dots X_1}{\dots X_2} \tag{7}$$

In Equation 7, relative behavior is a function of the relative rates (*R*), magnitude (*M*), and immediacy (the inverse of delay; *I*) of reinforcement, while *X* denotes any number of not listed dimensions (*X*) between two alternatives (1 and 2). Going one step further, raising each dimension to a sensitivity term grants Equation 8:

$$\frac{B_1}{B_2} = b * \left(\frac{R_1}{R_2}\right)^{S_R} * \left(\frac{M_1}{M_2}\right)^{S_M} * \left(\frac{I_1}{I_2}\right)^{S_I} * \left(\frac{\dots X_1}{\dots X_2}\right)^{S_X} \tag{8}$$

where $S_{R,M,I\dots X}$ represent scaled sensitives to changes in each of the relevant dimensions. Thus, matching is primarily a formalized restatement of empirical findings known to affect behavior (i.e., differential scaling of multiple dimensions). When coupled with modern statistical approaches (Young, 2017), Equation 8 is an effective tool at capturing individual differences and couches them in a theoretically parsimonious interpretation of differences in contact with relevant reward dimensions. Furthermore, although each dimension is said to be independent of others, the use of mathematical formulation would allow for empirical testing of such assumptions (Davison & McCarthy, 1988). Finally, for the sake of completeness, we formulate what has been an overriding assumption in Equation 9

$$\frac{B_1}{B_2} = \frac{V_1}{V_2} \tag{9}$$

where the relative rates of behavior (B_1 and B_2) are reflective of a relative value (V_1 and V_2) for available alternatives.

The progression of the matching equations exemplifies an approach to the treatment of data through quantitative models. Equation 5 is of particular importance, as R_e was meant to correct for some of the motoric problems previously noted with single-schedule measures of reward value. However, through formalizing testable predictions, it was shown that R_e could not fully reconcile confounding motoric effects. Such instances show how mathematical models can be "better than mere words" (Mazur, 2006, p. 276) because they require unambiguous specification of theoretical assumptions. Ultimately, single schedule measures would become less prevalent in the matching literature (but see Killeen, 1994), perhaps with the suggestion that they are less robust or representative measures of reward value. The prevalence of single schedules in substance abuse is also not necessarily due to failures of matching in drug-associated decisions, as they can describe drug and nondrug choices (Beckmann et al., 2019). Thus, the distinction of measurement is important for the quantification of value both in understanding factors that lead to drug choices and for the relatively new neuroeconomic field that seeks to find neural underpinnings of reward value in the brain (Glimcher et al., 2013).

Reward Value at Another Level: Neuroeconomic Reductionism

To this point, the discussion of reward value has revolved around the assumption of Equation 9: that measured behavior is a reflection of the hypothetical construct of value, or utility in the subjective case. Even in the matching equations, behavior is described *as if* a scaled multidimensional utility function is utilized to compare alternatives. The use of the phrase 'as if' is important (Glimcher, 2011; Glimcher, Dorris, & Bayer, 2005), as it serves to distance oneself from the false notion that individuals are consciously computing such values when they make decisions. However, if individuals do not consciously represent and adhere to their own utility functions, how are decisions made? The suggestion by the field of neuroeconomics is that decisions are made by the physiological architecture of neurons in the brain.

Glimcher (2011) suggested that one goal of neuroeconomics will be to move from the statement of individuals acting *as if* they represent values on a utility function to individuals making decisions *because* of value computations in the brain. However, for

brain value computations to be verified, understanding the implicated philosophical assumptions and a valid means of testing hypotheses are necessary. A core philosophical assumption for neuroeconomics is the idea of reductionism (McCauley, 2007), or that phenomena observed at one level of analysis can be reduced to another. For example, one mechanism as to why (non)humans may behave according to a monotonic utility function for food intake is because it leads to reproductive success. Thus, the concept of fitness maximization in evolutionary theory is a proposed mechanism for reducing the concept of rational human agency onto similar rational agency seen in animals (Kacelnik, 2006; Stephens, 2008). A further reduction could then imply that neuronal architectures lead to fitness maximization in animals that could be similarly present in mediating human decision making.

The second requirement for a neuroeconomic framework is a valid form of hypothesis testing. In order to test reductions in higher-order decision theories at the level of physiology, Glimcher (2011) suggested that researchers start by making assumptions from what is already known (i.e., utility models). Although no single utility function can describe choice data without fail, beginning with utility models allows for the identification of neural circuits involved in valuation when behavior is utility-theory compliant. Therefore, to test if brain areas "encode" reward value, physiological recordings of neuron activity or neurotransmitters are typically correlated with behavioral measures of reward value.

Mesocorticolimbic Encoding of Ostensible Reward Prediction Errors

One promising line of neuroeconomics research was the identification of an ostensible reward prediction error (RPE). Schultz and colleagues (Schultz, Apicella, & Ljungberg, 1993) measured ventral striatum neuron activation as thirsty monkeys learned a cued choice procedure in which a discriminative stimulus signaled operant-contingent availability of a delayed juice reward (see also Apicella, Ljungberg, Scarnati, & Schultz, 1991; Schultz, Apicella, Scarnati, & Ljungberg, 1992). Early in training, striatal neurons displayed a burst firing pattern of activation following reward delivery and a brief cessation of firing following reward omission. However, the pattern of neuronal activity also changed as the monkeys learned the task. Eventually the striatal neurons stopped firing when the reward was delivered and activated instead with the onset of the cue

signaling reward availability. The activity of the neurons therefore suggested not merely encoding a sensitivity to reward delivery, but also of cues providing information about reward and violations of reward expectations.

To further assess the plausibility of the RPE hypothesis, subsequent studies looked at neuronal activation under more uncertain conditions. For instance, providing reward in a random, behaviorally non-contingent manner can reestablish burst firing of striatal neurons to reward delivery (Mirenowicz & Schultz, 1994). Additionally, burst firing appears to scale with probability of reward. Fiorillo, Tobler, and Schultz (2003) trained monkeys on a Pavlovian procedure where a cue was probabilistically followed by a delayed reward. The results showed that burst firing to reward delivery was inversely related to the scheduled probability of that reward. However, burst firing to the cue predicting reward delivery was positively related to the scheduled probability, and activity between the cue and reward showed an inverted U-shaped pattern with maximal activity at 50% reward probability. Together, the results suggested that striatal neuron firing was encoding information about reward wherein reward-associated firing seemed contingent on how surprising an event was, while cue-associated activation was contingent on reward-predictive associability.

The finding of neural RPEs was exciting for multiple reasons, one being the possibility of an ostensibly common physiological signal for reward value (i.e., a common currency).The validation of a common currency mechanism would suggest that individuals do make decisions according to and because of neural representations of a utility function and provides a biological target to treat disorders stemming from reward pathologies (such as substance use disorders). Subsequent work has therefore attempted to evaluate the validity of the RPE as a common currency mechanism by testing commodities across a range of dimensions. For example, preclinical research has shown that RPEs in mesocorticolimbic areas appear to scale with changes in probability (Fiorillo et al., 2003; Hart, Rutledge, Glimcher, & Phillips, 2014; Lak, Stauffer, & Schultz, 2014), magnitude (Hart, Haney, Foltin, & Fischman, 2000; Kobayashi & Schultz, 2008; Padoa-Schioppa & Assad, 2006; Saddoris, Cacciapaglia, Wightman, & Carelli, 2015a) and delay (Day, Jones, Wightman, & Carelli, 2010; Kobayashi & Schultz, 2008; Roesch, Calu, & Schoenbaum, 2007; Saddoris et al., 2015b) for either rewards or reward-associated cues.

Additionally, with modest success, some studies have manipulated mesocorticolimbic activity to causally relate RPEs to preferences (Hamid et al., 2016; Saddoris et al., 2015b; Stopper, Maric, Montes, Wiedman, & Floresco, 2014).

Ostensible RPE signals have also been found in human mesocorticolimbic-associated nuclei using functional magnetic resonance imaging techniques (fMRI; Levy & Glimcher, 2012). fMRI measures blood oxygen level-dependent (BOLD) signals that act as a proxy for brain region activity (discussed more below in the *Oxygen Measurement and its Relation to Neuron Activity* section) and allow for recording of brain activity similar to electrodes used in preclinical studies. However, assessment of what reward-relevant dimensions human RPEs are sensitive to is somewhat precluded due to prevalent use of model-based fMRI (O'Doherty, Hampton, & Kim, 2007). In model-based fMRI, reward-relevant dimensions are often converted to a prediction derived from a utility function (such as an expected value) that is correlated with brain activity. As such, the scaling of area-specific activity with a singular dimension is less available. Still, several studies using magnitude, probability, and delay manipulations on choice alternatives have shown mesocorticolimbic activity is related to model-derived parameters (Abler, Walter, Erk, Kammerer, & Spitzer, 2006; Ballard & Knutson, 2009; Gläscher, Daw, Dayan, & O'Doherty, 2010; Kable & Glimcher, 2007; Knutson, Taylor, Kaufman, Peterson, & Glover, 2005), and suggests RPE scaling of reward-relevant dimensions. Additionally, two recent meta-analyses of fMRI studies assessing correlations between brain activity and decision-making have found homology of results to two mesocorticolimbic areas: the striatum (both ventral and dorsal) and prefrontal cortices (Garrison, Erdeniz, & Done, 2013; Levy & Glimcher, 2012). Thus, in order for greater homology between (pre)clinical literatures and more informed hypothesis testing, animal research should also increase its use of models describing preference behavior with neurological measures of reward value.

Molecular Analysis of Neurobehavioral Decision Making

One class of models that lends itself to the reductionist approach of neuroeconomics is reinforcement learning (RL) models (Sutton & Barto, 1998). RL models stem from Pavlovian conditioning research (Bush & Mosteller, 1953; Rescorla & Wagner, 1972) and suggest learning is dependent upon the unpredictability of rewards

and cues. The original intent of the Pavlovian equations was to model conditioned responses (CR) that occur when pairing a to-be conditioned stimulus (CS) with a biologically-relevant unconditioned stimulus (US; Rescorla & Wagner, 1972). The models later conceptually changed to modeling the associative strength between stimulus and reward in the modernized Equation 10

$$V_A^{t+1} = V_A^t + \alpha(\lambda - V_A^t) \tag{10}$$

where the associative strength (V) of stimulus A on the next trial (t) is a function of the difference between the current strength and asymptote of conditioning (λ) scaled by a learning rate parameter (α). Thus, similar to the idea of RPEs, the change in associative strength is a scaled correction between previous and new learning. Although inaccurate under certain conditions (e.g., Miller, Barnet, & Grahame, 1995), Equation 10 has served as a highly influential description of several conditioning phenomena (Miller et al., 1995; Rescorla, 2003).

Given the findings of Schultz and colleagues (e.g., Schultz et al., 1993), the use of RL models makes intuitive sense as it already applies a prediction correction algorithm as a learning mechanism. Additionally, application of RL models from computer science formulations (Sutton & Barto, 1998) and behavioral conditioning describe the neural RPE data well (Schultz, Dayan, & Montague, 1997). One formalization of how an RL model can update reward values is shown in Equation 11:

$$V_A^{t+1} = V_A^t + \delta_t \tag{11}$$

where the value of commodity A on the next trial (t) is defined by the summation of its current value and the reward prediction error (δ_t). The reward prediction error is defined in Equation 12.

$$\delta_A^t = \alpha(\lambda_A^t - V_A^t) \tag{12}$$

In Equation 12, the difference between a commodity's expected value (V) and the received reward (λ) is scaled by a learning parameter (α). Equation 12 is applied independently to a commodity when it is chosen to determine the expected value for each reward. Then, to determine action selection, the values of the commodities are compared according to a softmax decision rule (Sutton & Barto, 1998) to determine the probability of a choice:

$$pCh_A^t = \frac{1}{1 + exp(-[\beta(V_A^t - V_B^t)])} \quad (13)$$

In Equation 13, the probability of choosing commodity A on the current trial (*t*) is determined by the difference between reward values (*V*) for commodity A and B scaled by an inverse temperature parameter (β). Hierarchical RL models such as in Equations 11-13 have been shown to capture choice data of both humans (Rutledge et al., 2009) and animals (Groman, Rich, Smith, Lee, & Taylor, 2017). Additionally, RL models have the added benefit of being molecular analyses that capture choice-by-choice dynamics. While descriptive molar models such as the matching equation are often sufficient for describing choice behavior, their interpretation as a causal definition of choice behavior is strained due to its averaging over conditions of choice. Alternatively, RL models make a prediction for every choice that, if sufficient, grants one step closer to a causal interpretation.

The Value in Studying Value: Applications to Opioid Use Disorder

The previous sections illustrated how the determination of reward value is more complex than perhaps let on by single schedule measures prevalent in substance abuse research. In order to properly assess the conditions under which a commodity is favorably chosen, experiments need to explore a range of conditions across multiple dimensions of reinforcement relative to other commodities. The necessity for complex, relative models of reward also serves for a better representation of how drug users make drug-associated decisions in their natural environments. Substance users are not isolated to a box where the only available alternatives are to take drug or leisure activities (i.e., B_e). Thus, providing another biologically relevant commodity (e.g., food, money) is necessary for a better theoretical understanding of reward value and to improve the face validity of the model as a representative measure of what occurs in a natural environment (i.e., beyond the laboratory).

The discussion of valuation as it relates to substance use disorders is also predicated on one important point: that substance use disorders can be characterized as disorders of choice. Indeed, there is growing consensuses that choice mechanisms affect

illicit substance use. For example, the propensity to choose drug rewards can be altered by the available dose, cost to obtain drug reward, and the presence of an alternative reinforcer (Banks & Negus, 2012; Bickel, Johnson, Koffarnus, MacKillop, & Murphy, 2014; Davis et al., 2016; Lamb et al., 2016; Moeller & Stoops, 2015). Such results therefore suggest that drug-associated decision making is similarly economic as non-drug choices.

A competing hypothesis with drug taking as a disorder of choice is that of compulsion (Kalivas et al., 2005; Lubman, Yücel, & Pantelis, 2004; Vandaele & Janak, 2018). For a compulsive drug user, drug use is poorly mediated (if at all) by value-based decision making. Ideas of compulsive drug use are typically predicated on the idea that continued drug use leads to brain changes (e.g., Everitt & Robbins, 2016; Volkow, Koob, & McLellan, 2016; Volkow, Wang, Fowler, & Tomasi, 2012) that lead to an insensitivity to behavioral consequences. However, multiple theoretical and empirical problems exist with the idea of compulsion, not the least of which is that the definition of compulsion changes over time (Heyman, 2003, 2013a; Hogarth, 2018). For example, the brain changes that compulsive use is predicated on involve post-hoc correlational evidence that lacks causal inference. Additionally, incidents of drug experimentation leading to regular use is low, contingency management programs offering incentives contingent on drug-abstinence are highly effective at curbing long time use, and users often remit without treatment (Davis et al., 2016; Heyman, 2003, 2013b). Even rats given extended access to self-administer drugs of abuse often do not show potentiated drug preferences after they become labeled “compulsive users” (Caprioli et al., 2015; Lenoir et al., 2007; Schwartz et al., 2017). Thus, the suggestion is not that brain changes do not occur, but that the changes may not necessarily preclude an individual from making value-based economic decisions (Heyman, 2013b).

Although drug-associated brain changes may not be a useful determinant for describing drug preferences, they may serve as a useful testing ground for neuroeconomic hypotheses. At present, the idea of a cardinal utility scale in the brain is predicated on mesocorticolimbic activity scaling with different metrics of behaviorally derived reward value. However, drugs of abuse also directly alter mesocorticolimbic activity through their pharmacodynamics (Iversen, Iversen, Bloom, & Roth, 2008; Pierce & Kumaresan,

2006) and produce activity that may not necessarily be related to value. For instance, sensitization or tolerance to the pharmacodynamics of opioid μ receptor agonists (Fields & Margolis, 2015; Grecksch et al., 2006; Williams, Christie, & Manzoni, 2001) may produce signals within mesocorticolimbic areas that do not scale with behavioral reward value measures. Dissociation between brain and behavioral measures to reward delivery would subsequently require a careful reconsideration of neuroeconomic hypotheses. To date, (pre)clinical studies have shown that mesocorticolimbic areas, such as the striatum, are activated by cocaine and heroin reward in a similar manner to food rewards (Cameron, Wightman, & Carelli, 2014; D'Souza & Duvauchelle, 2008; Guillem, Brenot, Durand, & Ahmed, 2018; Phillips, Stuber, Heien, Wightman, & Carelli, 2003; Risinger et al., 2005; Saddoris, Wang, Sugam, & Carelli, 2016), but studies have not yet scaled brain activity with reward-value measures to test neuroeconomic hypotheses.

Invoking ideas of compulsion may also not be necessary to describe dysregulated drug use, as it is possible that such continued use actually stems from valuation mechanisms (Heyman, 2013b). Rather than drug use being a habit separate from value-based decisions, a valuation mechanism would suggest substance abuse is a behavior allocation disorder (Lamb & Ginsburg, 2018), comprised of competing contingencies in what commodities produce greater local versus global utility (Herrnstein & Prelec, 1991; Rachlin, 1997, 2000). For instance, melioration theory suggests drug commodities offer relatively higher immediate utility than nondrug commodities. Nondrug commodities may offer more utility in the long term (such as hard work leading to career advancement or avoidance of job loss), but these consequences are discounted because they are delayed.

The above depiction of a drug user may appear similar to compulsion theories on the surface to some, but there are important differences in what mediates regular use. For a compulsive user, regular use stems from a lack of control where the user is not making economic decisions and is a static agent. However, the user under melioration theory is not compulsive and economic decisions are still being made on a local expected utility basis. The distinction of maintaining a user as an economic agent may seem trivial, but it has important implications, such as changing the prognosis of treatment from being necessarily pharmacological to potentially behavioral. Additionally, if drug users do

maintain economic agency, then understanding how users adapt to changing conditions may provide new therapeutic targets to adapt behavior away from drug preferences.

Current Dissertation

Dynamic Decisions Across Multiple Dimensions of Reward Value

In order to understand opioid-related decision making, research from behavioral economics and neuroeconomics has suggested that experiments need to account for the contextually dependent and relational nature of reward value. Furthermore, experiments using single-schedule measures common in substance abuse research often do not model decision making in these more complex environments. The current dissertation therefore characterized the degree to which opioid-related decision making operates according to a value-based mechanism through economic drug choice procedures.

Three experiments modeled value-based decision making in rodents based on a procedure previously used with humans (Rutledge et al., 2009). Experiment 1 validated the procedure with drug-naïve rats who chose between two probabilistically delivered food rewards. The reward probabilities of both choices changed in a dynamic, unpredictable manner, such that maximizing reward intake required constant updating of reward value based on trial-outcome feedback. Additionally, given that evidence that reward-associated cues can alter reward value (e.g., Barrus, Cherkasova, & Winstanley, 2015; Zentall & Stagner, 2011), the rats were split between Signaled and Unsignaled groups. Upon making a choice for either alternative, the Unsignaled group received a choice-dependent cue that was uninformative of trial outcome; this group served as a control to represent more commonly used procedures where choices either result in food reward or food omission without any cues. Alternatively, the Signaled group received differential reward-associated cues (Smith, Hofford, Zentall, & Beckmann, 2018). Choice of one option resulted in an uninformative cue similar to the Unsignaled group. However, choice of the alternative option produced distinct win- and loss-paired cues that were always followed by reward or omission, respectively. Following sufficient learning of task contingencies, the reward magnitude was increased for one counterbalanced option in the Unsignaled group and the option with reward-associated cues in the Signaled

group. Choice data were then analyzed using molar matching and molecular RL models to assess economic parameters.

Experiment 2 trained a separate cohort of rats on a probabilistic allomorphic drug versus food choice by replacing the food reward for one counterbalanced option in the Unsignaled group and the option with reward-associated cues in the Signaled group with a self-administered, intravenous infusion of remifentanil (REMI). REMI is a potent μ opioid receptor agonist (Crespo, Sturm, Saria, & Zernig, 2005) similar to morphine and heroin (Fields & Margolis, 2015). Three reward magnitudes (i.e., doses) were tested similar to the three food reward magnitude manipulations in Experiment 1. Finally, Experiment 3 extended the REMI choice model by recording brain activity in the nucleus accumbens (NAc), a prominent mesocorticolimbic nucleus in the ventral striatum that is broadly associated with reward-associated decision making (Salgado & Kaplitt, 2015). NAc activity was recorded using oxygen-sensitive (O_2) microelectrode arrays (MEAs) implanted into the NAc. The use of O_2 increases translational fidelity with fMRI BOLD measures used in clinical research through a common measure of brain activity. However, O_2 is recognized as a more general measure of activity than other available alternatives.

Oxygen Measurement and its Relation to Neuron Activity

Oxygen as a measure of neuronal activation is based on cerebral metabolism of O_2 by local cells as well as the cerebral blood flow/volume in the measured area. In fMRI BOLD signals, oxidative metabolism occurs in a local area that deoxygenizes hemoglobin. The loss of O_2 from hemoglobin then causes co-localized iron to become paramagnetic and produces a measurable effect on the surrounding magnetic field of water molecules (Attwell & Iadecola, 2002). However, increases in cerebral blood flow/volume can exceed deoxygenation in activated areas that often leads to an increase in net O_2 (Ekstrom, 2010).

It has been estimated in primates and rodents that a large proportion of oxygen changes are related to neuronal activity (Attwell & Laughlin, 2001), but exact activity specification has not been determined. While spiking is related to cerebral oxygen responses, post-synaptic events and neurotransmitter release are also thought to contribute as O_2 responses can exceed electrophysiological or local field potential

measures of neuronal spiking (Attwell & Iadecola, 2002; Buxton, 2012; Ekstrom, 2010; Sheth et al., 2004). Thus, the current paper will use the general term of "*NAc activity*." Despite being somewhat vacuous, the term NAc activity avoids potential over-attribution of any one physiological mechanism, save those that are neuronally mediated.

Measuring Oxygen with Microelectrode Arrays

Although rodent fMRI is possible, the necessary immobility of the animal precludes any ability for behavioral assessment. As an alternative, local oxygen dynamics can be captured using oxygen-sensitive electrodes in freely moving animals (Lowry et al., 2010). Electrode amperometry measures of O_2 work by applying a constant potential of -0.6 (vs. reference) that leads to O_2 reduction at the surface of the MEA (Li, Schwarz, & Gilmour, 2015). The reduction occurs in two steps:

$$O_2 + 2H^+ + 2e^- \rightarrow H_2O_2$$

$$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O$$

and the obtained measure is a change in current induced by electrons transferring from the MEA to oxygen. Amperometry measures of O_2 have high spatial (μm) and temporal (100 Hz) resolution (Ledo et al., 2017) similar to electrophysiology or voltammetry measures used in preclinical research while also maintaining an O_2-dependent measure akin to fMRI BOLD signals. Electrochemistry measures of O_2 and BOLD are also highly correlated (Lowry et al., 2010). When compared to rodent fMRI, the time course of fMRI O_2 changes correlates with 99% of variance accounted for by amperometry. Additionally, amperometry-derived O_2 signals show similar functional responses to reward delivery in the NAc as fMRI studies (Francois, Conway, Lowry, Tricklebank, & Gilmour, 2012; Francois et al., 2014). Thus, the O_2 sensitive electrode recordings can be regarded as a proxy of fMRI BOLD signals.

An Improved Model of Drug-Associated Value

As previously described, models of substance abuse can be improved by accounting for the multidimensional and relative nature of a commodity's reward value. As such, the current experiments sought to improve upon traditional measures of drug-associated reward value in several ways. First, all measures of reward value were a relative choice measure. That is, two commodities were always concurrently available that allowed for a comparison of preference for one option (e.g., Option A) relative to an

objectively defined alternative (i.e., Option B). Second, reward value was assessed in more complex, dynamic environments than have historically been used in substance abuse research. Finally, brain activity was recorded using an O_2-sensitive measure that increased translational fidelity with clinical BOLD measures and was compared to multiple models of reward value. Together, the present improvements allowed for a more explicit assessment of whether REMI-associated choices are mediated by value-based or compulsive decision-making mechanisms.

CHAPTER 2:
DYNAMIC DECISION-MAKING FOR ISOMORPHIC FOOD VERSUS FOOD REWARDS

Experiment 1

In the current experiment, decision dynamics were characterized using 'natural' rewards (i.e., food) in a probabilistic choice procedure. In doing so, the results of subsequent experiments assessing drug-associated decision making can be compared back to the current experiment to avoid misattribution of any potential effects as being unique to drugs of abuse.

Methods

Subjects

Twelve adult male Sprague-Dawley Rats (Harlan Inc; Indianapolis, IN, USA) were used in the experiment. Rats were individually housed on a 12:12 hour light:dark cycle (lights on at 7:00 a.m.) and had free access to water but were food restricted to 15 g of standard lab chow per day (Harlan Inc.) post session. All research was approved by the University of Kentucky Institutional Animal Care and Use Committee (Protocol #2011-0885).

Apparatus

Experiments were conducted in Med Associates (St. Albans, VT) conditioning chambers (ENV-008) within a sound-attenuating box (ENV-018). On the front panel inside the chamber were two white jewel lights (ENV-221M) above two retractable response levers (ENV-122CM) flanking a magazine receptacle (ENV-200R2MA) equipped with a head entry detector (ENV-254-CB) that received 45-mg sucrose pellets (BioServ Precision Pellets, Flemington, NJ) from a pellet dispenser (ENV-203M-45). Directly opposite the levers on the back panel were two nosepoke response receptacles (ENV-114BM) and a house light (ENV-227M) centrally positioned at the top of the back panel. All chambers were connected to a personal computer running MED-PC IV (Tatham & Zurn, 1989).

Establishing operations

Prior to the experiment proper, rats were pre-trained on four consecutive procedures: magazine shaping, nosepoke training, nosepoke training with terminal link stimuli, and orienting responses. During magazine training, rats were placed in the chamber under a fixed-time (FT) 60-s procedure where a pellet was dropped into the magazine every 60 s concurrent with the illumination of the house-light for 5 s. Sessions lasted for 1 hr but only 30 pellets were delivered into the chamber. After all rats ate all pellets for two consecutive days, the rats were trained to make nosepoke responses. Sessions began with the illumination of either the left or right nosepoke and was denoted the initial link of the eventually chained stimuli. A response to the nosepoke would then offset the receptacle and deliver a food pellet to the magazine. Nosepoke training consisted of 30 trials (15 left responses, 15 right responses) with the order of left versus right nosepoke presentation pseudorandomized across trials such that no nosepoke appeared more than two consecutive trials.

Following two days of completing all trials during nosepoke training, terminal link stimuli were added to the procedure. During terminal link training, responding to a lit nosepoke (i.e., the initial link) resulted in the presentation of either the left or right lever on the front panel for 10 s (i.e., the terminal link), followed by lever retraction and the delivery of 1 food pellet to the magazine. The spatial counterbalancing of the nosepoke–lever combinations was consistent to one side of the chamber from the animal's perspective; for example, a response to the left nosepoke on the back panel produced the right lever on the front panel, and vice versa for the other alternative. Here, the lever is denoted as the terminal link because it is the final stimulus presented prior to reward delivery. After completing all trials for two sessions, orienting-response training began. During orienting-response training, illumination of the house-light signaled the beginning of a trial that required a head entry into the magazine to turn off the house-light. Following an orienting response (i.e., the head entry to turn off the house-light), either the left or right nosepoke illuminated in a pseudorandom order, and the procedure continued as previously described. Orienting-response training continued until all trials were completed for two days.

Isomorphic decision-making procedure

After all rats completed the establishing operations, training began on a dynamic probabilistic decision-making procedure illustrated in Figure 2.1 (Rutledge et al., 2009). Sessions consisted of four, 30-trial blocks (120 trials total). Each block was characterized by different relative reward probabilities. The possible relative reward rates, expressed as Option A:B, were 6:1, 2:1, 1:2, and 1:6 (see Figure 2.1c), and both Option A and B delivered 1 food pellet for a win trial. As an example, in the 6:1 block, Option A rewarded choices six times more often than Option B. The reward schedules were randomized such that the start of the session could begin with any one of the four reward schedules (i.e., 6:1, 2:1, 1:2, 1:6). After the first block, the determination of remaining blocks was pseudorandomly determined without repeat such that the choice alternative with the higher chance of reward (i.e., the "richer" option) switched to the opposite choice alternative. For instance, if the first block was the 6:1 condition, the second block was required to be either the 1:2 or 1:6 conditions.

The rats were also split into the Signaled and Unsignaled groups. As shown in Figure 2.1, the sequence of events following choice of either Option A or B was group dependent. For the Unsignaled group, choice of either option resulted in the presentation of an associated lever stimulus for 10 s followed by the probabilistic delivery of 1 food pellet according to the current reward schedule. For the Signaled group, choice of Option B was the same as the Unsignaled group. However, choice of Option A produced a lever stimulus *only* when a reward was to be delivered (i.e., a win); a white jewel light illuminated to signal an upcoming loss. The overall reward probabilities for the two groups were identical. Thus, the only difference between groups was that choices of Option A for the Signaled group produced a stimulus that was informative of the forthcoming trial outcome 10 s later, while Option B for the Signaled group and Options A and B for the Unsignaled group produced stimuli that were not informative of the upcoming trial outcome. After reward delivery or omission, trials were separated by a 10-s intertrial interval (ITI) during which all stimuli were off. Training continued until all rats significantly decreased their choice of Option A as its relative reward rates declined for fifteen session. Training lasted for 70 sessions.

Reward magnitude manipulations

Following initial training, the reward magnitude associated with Option A was subsequently increased. For the duration of the reward magnitude manipulations, Option B continued to deliver one sucrose pellet on win trials. However, Option A increased in reward magnitude to two and three sucrose pellets in Phases 2 and 3, which each lasted for 10 sessions for both groups.

Data Analysis

Molar choice data

Molar choice data were calculated as the proportion of choices for Option A as a function of Option A's relative reward rates averaged over the last five days of training for each reward magnitude condition. Molar data were then analyzed using three non-linear derivations of the generalized concatenated matching equation (Equation 7) with nonlinear mixed effects (NLME; Smith et al., 2018; Young, Clark, Goffus, & Hoane, 2009) models in the NLME package in R (Pinheiro, Bates, DebRoy, & Team, 2016). For the behavioral data where both options probabilistically rewarded one food pellet, Equation 14 was used and is referred to as the Generalized Matching equation:

$$\frac{B_A}{B_A + B_B} = \frac{1}{1 + \left(\frac{R_A}{R_B}\right)^{S_R} * b} \tag{14}$$

In Equation 14, B_A and B_B represent choices for Option A and B, expressed as a function of the relative reward rates for Option A (R_A) and B (R_B) multiplied by a bias term (b). The bias term (b) captures preferences for Option A or B irrespective of the relative reward rates between the two options. Lower values of b reflect greater bias for Option A; values close to 1 reflect no bias. S_R captures individuals' sensitivities to changes in the relative reward rates between Options A and B. Perfect matching between behavior and relative reward rates would reduce the value of S_R to 1.

To account for the later magnitude manipulations, an additional term was added to the Generalized Matching equation (Equation 14) to capture changes in choice due to the changes in reward magnitude. Specifically, Equation 15, referred to as the 2-Dimensional (2-D) Generalized Matching equation, concatenated the relative reward magnitudes of Options A and B.

$$\frac{B_A}{B_A + B_B} = \frac{1}{1 + \left(\frac{R_A}{R_B}\right)^{S_R} * \left(\frac{M_A}{M_B}\right)^{S_M}} \quad (15)$$

In Equation 15, the relative reward magnitudes for Option A (M_A) and B (M_B) are raised to a sensitivity term, S_M, that scales how sensitive individuals are to changes in the relative reward magnitudes.

While Equation 15 does well to describe choice data across individuals choosing under similar conditions, the current experiment employed group-dependent use of reward-associated cues not necessarily associated with the dimensions of reward rate or magnitude. Therefore, should any cue-dependent effects emerge, it is empirically undetermined how to best quantitatively describe the effects. Some research has suggested potential cue effects are mediated by enhancing the perceived reward magnitude of choice alternatives with win-paired cues (i.e., Option A for the Signaled Group; Smith et al., 2017a; Smith et al., 2018), while others have suggested that choice alternatives with cues gain value via other dimensions not manipulated here (Cunningham & Shahan, 2019; Ojeda, Murphy, & Kacelnik, 2018; Trujano, López, Rojas-Leguizamón, & Orduña, 2016). If potential cue effects are mediated by changes in subjective reward magnitude, group differences should be captured by the relative reward magnitude scaling parameter, S_M. However, in the case that changes in perceived reward magnitude are not mediating potential group effects, Equation 16 (2-D Exchange Matching) can scale reward value differences in terms of Option A and B's relative substitutability (Rachlin et al., 1976).

$$\frac{B_A}{B_A + B_B} = \frac{1}{1 + \left(\frac{R_A}{R_B}\right)^{S_R} * \left(\frac{M_A}{Ex}\right)^{S_M}} \quad (16)$$

In Equation 16, the previous term for Option B's reward magnitude is replaced by the scaling constant, *Ex*. *Ex* acts an exchange rate for the subjective value of Option B relative to Option A in units of sucrose pellets. For instance, when the reward magnitude of Option A is 1, an *Ex* value of 1 suggests the two options have equivalent reward magnitudes and are equally substitutable for each other. Values of *Ex* less than 1 suggest that Option B is worth less than Option A.

For all models, Subject was entered in as a nominal random factor, Group as a nominal, between-subject fixed factor using effects coding, and sensitivity parameters (i.e., S_R and S_M) were fit across their relevant dimensions (i.e., relative reward rates and magnitudes). In order to determine the best-fitting model, model comparisons were conducted using Akaike Information Criterion (AIC; Symonds & Moussalli, 2011; Wagenmakers & Farrell, 2004) where smaller values indicate better model fits. When comparing AIC values between a current and candidate model, changes in AIC values (ΔAIC) ≥ 0 reflect evidence favoring the current model, while $0 \geq \Delta AIC \geq -2$ reflect weak evidence, $-4 \leq \Delta AIC \leq -7$ reflect moderate evidence, and $\Delta AIC \leq -10$ reflect substantial evidence for the candidate model. For the current model comparisons, an absolute ΔAIC of at least 4 was required to select a more complicated candidate model over a simpler one with fewer free parameters. All parameter estimates, that are negative in the raw form, are shown as absolute values to aid interpretation.

Reinforcement learning models

In order to test if the results from the molar matching models were applicable at the molecular level, RL models were fit to characterize the same data as the matching models at a trial-by-trial resolution. A basic RL model was previously described in Equations 11-13, but, as Equation 13 was modified to Equation 17 for the current experiment, the equations are repeated here for convenience:

$$V_A^{t+1} = V_A^t + \delta_A^t \tag{11}$$

$$\delta_A^t = (\lambda_A^t - V_A^t) \tag{12}$$

$$pCh_A^t = \frac{1}{1 + exp(-[\beta(V_A^t - V_B^t)] + c[Ch_A^{t-1} - Ch_B^{t-1}])} \tag{17}$$

In the valuation Equations 11 and 12, the value of Option A (V_A), on a given trial, *t*, is determined by its current value plus the reward prediction error, δ, or the difference between the obtained (λ_A) and expected (V_A) reward scaled by the learning rate parameter, α. Equation 17 then determines the probability of choosing Option A through a softmax decision rule by scaling the relative value difference of Options A and B according to the inverse temperature parameter, β. Equation 17 also has the free parameter *c* that weights an individual's tendency to repeat the choice made (*Ch*) on the

previous trial (*t-1*) independently of reward value. Ch_A equals 1 when the previous choice was for Option A (i.e., $Ch_{B,t-1} = 0$), and Ch_B equals 1 when the previous choice was for Option B (i.e., $Ch_{A,t-1} = 0$). Model comparisons began by fitting the Base RL model as formulated in Equations 11, 12, and 17. The ability of RL models to parameterize rats' choices was also compared to a Chance model that predicted random choices (i.e., probability of choosing A = 0.5).

In order to test if scaling reward magnitude in a similar way as the matching equations is warranted, the reward prediction error from Equation 12 in a subsequent Scaled Single-Learning model was modified to Equations 18 and 19:

$$\delta_A^t = ((\lambda_A^t/\lambda_B^t)^{S_M} - V_A^t) \tag{18}$$

$$\delta_B^t = ((\lambda_B^t/\lambda_A^t)^{S_M} - V_B^t) \tag{19}$$

where the obtained reward (λ) for Option A or B is expressed as a ratio relative to the other option and then raised to the sensitivity to reward magnitude parameter, S_M. Although RL models have traditionally assumed that choices are made by comparing relative values, valuation functions such as Equations 11 and 12 typically ascribe value to a commodity independently of others. In Equations 18 and 19, value is instead also said to be relative to and scaled with other available commodities. For the current experiment, λ_A represents the magnitude in units of sucrose pellets for Option A on win trials, while λ_B represents either the magnitude for Option B or the exchange rate, *Ex*, depending on the outcome of model comparisons from the molar matching models. Conversely, the ratio of λ_A and λ_B is recoded as zero on loss trials.

Subsequent RL models were also assessed that included additional α parameters dependent on the choice made (Scaled Double-Learning (Option)), the choice's outcome (Scaled Double-Learning (Outcome)), or the interaction between choice and outcome (Scaled Quad-Learning (Option×Outcome)). For instance, given that Option A and B represent two different signaling conditions for the Signaled group, the weighting of trial outcomes may not be uniform across choice options. Thus, the three previously mentioned models were evaluated using model comparison methods that allowed α values to vary as a function of Options A and B (i.e., α_A, α_B), choice outcome (i.e., wins vs. losses; α_{Win}, α_{Loss}), and wins and losses dependent on the chosen option (α_{AWin}, α_{ALoss},

α_{BWin}, α_{BLoss}). The *fmincon* optimization algorithm in MATLAB was used to fit each RL model to individual rat's trial-by-trial choice data using maximum likelihood estimation (Myung, 2003). Model fitting occurred by randomly selecting 100 uniformly distributed initial starting values constrained on the following intervals: $\alpha \in [0, 1]$, $\beta \in [0, 10]$, $c \in [-1, 1]$, and $S_M \in [0, 5]$. To quantify potential differences among parameters, group effects were assessed using a series of Wilcoxon rank-sum tests (coded as Signaled-Unsignaled). In the case that the best fitting model included more than one α, statistical differences across α parameters were assessed using pairwise Wilcoxon signed-rank tests with Hochberg corrections for multiple comparisons (Hochberg, 1988). Finally, RPEs combining all α, group-dependent λ, and S_M parameters were calculated when V_A and V_B equaled zero as a visual aid to illustrate potential RPE differences across groups.

Results

Molar analysis

Figure 2.2a shows the average proportion of choices for Option A as a function of Option A's relative reward rates when the reward magnitude for both options was one food pellet. Overall, both groups tended to prefer the option providing greater reward odds (or the "'richer" option), with preferences being stronger the more disparate the relative reward rates. In order to quantify the effects of dynamically changing reward probabilities on choice of Option A, a generalized matching nonlinear mixed-effects model (NLME; Equation 14) was fit to the data. The model captured the data well, and results showed that sensitivity to reward rates was significantly higher than zero [$F(1, 33) = 82.18$, $p < .001$]. Both the Signaled [$S_R = 0.31$] and Unsignaled [$S_R = 0.29$] groups showed comparable [$p = .834$] decreases in preference for Option A as its relative reward rate declined. However, as indicated by the parameter values being less than 1, both groups displayed somewhat shallow decreases in choices relative to the scheduled changes in relative reward rates. Additionally, the Signaled group [$b = 0.65$] showed significantly more bias for Option A than the Unsignaled group [$b = 1.05$], that showed very little bias for either option [$F(1, 33) = 9.31$, $p = .005$].

Subsequent increases in Option A's reward magnitude also showed that choices tended to favor Option A's increased magnitude (see Figure 2.2b-c). Model comparisons between different forms of the generalized matching equations are shown in Table 2.1.

Interestingly, the 2-D Matching Model showed relatively poor fits to the data (not shown) and was unable to effectively capture the subjects' data. A global (i.e., equivalent across groups) exchange rate parameter (*Ex*) was then added to scale Option B's reward magnitude in terms of Option A in the 2-D Global Exchange Matching model. The inclusion of *Ex* was beneficial as the AIC value of the 2-D Global Exchange Matching Model was -17.39 units lower, suggesting substantial evidence in favor of including an exchange parameter. Finally, when the exchange rate parameter was allowed to vary for both groups in the 2-D Exchange Matching model, model comparisons again substantially favored the group-dependent exchange parameter. Overall, the model comparison results strongly favored selection of the matching models that included a group-dependent exchange rate parameter.

The absolute value parameter estimates for the final model are shown in Figure 2.3. Similar to initial training, sensitivity to reward rates (i.e., S_R) was significantly greater than 0 [$F(1, 127) = 213.26$, $p < .001$] due to decreasing relative reward rates producing decreases in Option A preferences that were comparable across groups [$p = .405$]. Increasing the reward magnitude for Option A also revealed that sensitivity to reward magnitude (i.e., S_M) was significantly greater than 0 [$F(1, 127) = 194.10$, $p < .001$] with a significant effect of Group [$F1, 127) = 13.90$, $p < .001$] due to the Unsignaled group [$S_M = 1.06$)] showing significantly greater sensitivity to the changes in reward magnitude than the Signaled group [$S_M = 0.62$]. Finally, there was a significant effect of Group on exchange rates [$F(1, 127) = 10.13$, $p = .002$] due to the Unsignaled group showing comparable exchange between Options A and B [$Ex = 1.08$], whereas an unsignaled pellet did not substitute as well for a signaled pellet in the Signaled group [$Ex = 0.54$]. Finally, as an example of the scaled differences in reward magnitude between groups, Figure 2.2d shows the Signaled groups proportion of choices for Option A when the reward magnitude equaled 1 pellet and the Unsignaled group's proportion of choices for Option A when Option A's reward magnitude equaled 2 pellets. By doubling the reward magnitude of Option A for the Unsignaled group, the Unsignaled group's proportion of choices for Option A approximated the Signaled group's choice function when Option A only rewarded 1 pellet.

Molecular analysis

The molar analyses suggested that Option A's reward-associated cues for the Signaled group produced significant alterations in both sensitivity to magnitude and the exchange rate parameter. If the scaling of the subjective reward magnitude between Option A and B was indeed a determinant of choice in the current experiment, then similar effects should appear in the molecular trial-by-trial RL model analysis. Consistent with the molar results, adding group-dependent exchange rates for Option B's reward magnitude from the molar matching model as well as a sensitivity to magnitude parameter in the Scaled Single-Learning model substantially improved overall fits relative to the Base RL model (overall ΔAIC = -221.73; see Table 2.2). Adding additional α learning rate parameters also sufficiently improve model fits when they were Option-dependent or Outcome-dependent. But, including learning rate parameters that were both Outcome- and Option-dependent in the Scaled Quad-Learning model was the overall best fitting model for both groups (Table 2.2).

Figure 2.4a-c shows three example model fits that approximate the average AIC (± 1 standard deviation). Overall, the RL model successfully parameterized the subjects' trial-by-trial choices [ΔAIC = -684.42 relative to a chance model] across four levels of relative reward rates, three levels of relative reward magnitudes, and two levels of reward-associated stimuli. The parameters from the Scaled Quad-Learning model are shown in Figure 2.5a. The Signaled group showed (1) significantly higher β values [Z = 2.01, p = .028] indicating a greater tendency to exploit the Option with greater model-derived reward value, and (2) significantly higher α values for Option A losses [Z = 2.20, p = .028]. However, both groups scaled the changes in reward magnitude in a similar way and perseverated on choices a similar amount. For within-group tests of the α parameters, no significant differences were present after correcting for multiple comparisons. Finally, Figure 2.5b illustrates the calculated RPEs from the Scaled Quad-Learning model. As suggested by the AIC-warranted group-dependent exchange parameter, the Signaled group ordinally showed greater updating of reward value for Option A relative to the Unsignaled group that showed ordinally greater updating of reward value for Option B.

Discussion

The present experiment sought to establish the dynamic decision-making model using natural rewards to allow for better interpretation when decisions are made between allomorphic drug vs. food choices. To that end, this experiment showed that rats engaging in a randomized probabilistic decision-making procedure could effectively track the changes in relative reward rates and allocate their decisions appropriately, similar to monkeys (Lau & Glimcher, 2005) and humans (Rutledge et al., 2009). As indicated by the significant sensitivity to relative reward rates parameter, rats in both groups favored Option A when the relative reward rates also favored Option A and then gradually shifted preferences concurrent with shifting reward rates favoring Option B. Sensitivity to relative reward rates also showed evident undermatching (i.e., values less than 1), suggesting possible difficulty in discriminating the randomized relative reward rates. Indeed, probabilistic decision-making procedures that use consistent ascending or descending changes in reward probability (e.g., Smith et al., 2018; Yates et al., 2016) have shown qualitatively sharper preference changes as a function of reward probability changes. However, both groups in the current experiment still showed a nearly 30% decrease in preference between the 6:1 and 1:6 relative reward rate conditions, suggesting the tracking of and appropriate behavioral allocation to relative reward rates.

In addition to changes in relative reward rates shifting choice preferences, changes in relative reward magnitudes also significantly altered choice patterns. Specifically, the increase of reward magnitude associated with Option A from 1 to 3 pellets also resulted in an increased preference for Option A similar to previous experiments (Landon, Davison, & Elliffe, 2003; Marshall, Smith, & Kirkpatrick, 2014; Neuringer, 1967). Furthermore, the ability of Equation 16, which is an extension of the generalized concatenated matching equation (Baum, 1974; Rachlin, 1971), to effectively describe choices provides further evidence that relative reward rates and magnitude can produce independent effects on behavior (Kyonka, 2008; McLean & Blampied, 2001). That dimensions of reward value can produce independent effects on choice complicates the use of other unidimensional scales such as unit price (a ratio of magnitude divided by cost), as it suggests asymmetric effects when either side of the ratio is manipulated (Madden et al., 2000; Smith et al., 2016).

While choices within groups was well characterized, model comparisons revealed that choice preferences across the Signaled and Unsignaled groups were not accounted for by sensitivity to changes in either relative reward rates or magnitudes. The inability of the S_M parameter to capture group differences is particularly notable, as previous research has characterized the effect of reward-associated cues as altering a choice alternative's perceived reward magnitude (e.g., Smith et al., 2017a; Smith et al., 2018). However, previous studies suggesting alterations in reward-magnitude sensitivities only used a single relative magnitude condition similar to Figure 2.2a. Under such conditions, the parallel shift in preferences for Option A by the Signaled group is suggestive of a scalar increase in reward value for that choice alternative such as through a reward magnitude mechanism. It was not until the current experiment where the assumption of altered reward magnitude sensitivities, tested through multiple relative reward magnitude conditions, was coupled with model comparisons so that the hypothesis could be thoroughly evaluated. The present experiment therefore serves as an exemplar for the previous statement that models can be "better than mere words" (Mazur, 2006, p. 276), as the conceptualization of reward-associated cues producing a scalar increase in reward value is warranted, albeit not via a sensitivity to reward value mechanism as ascribed in the generalized matching equations.

Seemingly contrary to the model comparison results, the 2-D Exchange Matching model also showed a significant group effect on sensitivity to magnitude; this effect, which only appeared in the context of including an exchange rate parameter (*Ex*), was not expected and likely did not account for the group differences in other contexts. If differences in sensitivity to magnitude were the primary determinant of choice differences between the groups, the 2-D Generalized Matching equation would have produced better AIC values. Rather, the effect seems to reflect a threshold in how high rats' absolute preferences for Option A in both groups can be, which was reached earlier in the Signaled group due to their starting at a higher preference for Option A at baseline (i.e., the 1 vs. 1 food pellet condition.). Additionally, the group effect on S_M does not necessarily reflect a violation of matching. Matching assumes that the scaling of a dimension within a particular context is invariant relative to models which allow differential scaling parameters for each (e.g., Davison & McCarthy, 1988; Kyonka,

2008). Stated differently, matching assumes that only one S_R value is necessary for all relative reward rate manipulations. In the present experiment, a single S_R and S_M parameter was shown to provide excellent fits to the data, suggesting that the group effect on S_M likely reflected that sensitivities to a reward dimension were context-dependent, but invariant within one context.

Other research groups have attempted to characterize similar cue effects to that seen here as being driven by mechanisms of either other reward value dimensions or cognitive mechanisms. For instance, the temporal information model suggests that the reason the Signaled group's reward-associated cues produced preference bias for Option A was due to the value of having information about forthcoming rewards temporally sooner than those of choices of Option B (Cunningham & Shahan, 2019; Cunningham & Shahan, 2018). Alternatively, other models have suggested mechanisms less scalable in terms of reward dimensionality, such as the reward-associated cues producing a positive state change between uncertainty to certainty about reward (Case & Zentall, 2018; McDevitt, Dunn, Spetch, & Ludvig, 2016). Although the present experiment cannot discern these proposed mechanisms, the phenomenon of reward-associated cues biasing choice is robust and present across a number of species (Blanchard, Hayden, & Bromberg-Martin, 2015; Iigaya, Story, Kurth-Nelson, Dolan, & Dayan, 2016; Molet et al., 2012; Smith et al., 2018; Smith, Beran, & Young, 2017b; Vasconcelos, Monteiro, & Kacelnik, 2015; Zentall, 2016).

Equation 16 offers another way to quantitatively describe the effect of reward-associated cues in terms of substitutability (Green & Freed, 1993) through the exchange rate parameter. Substitution is defined as the degree to which consumption of an alternative commodity changes as the cost or magnitude of the present commodity changes. For the present experiment, as the relative reward rates stopped favoring Option A, choices of Option B increased for both groups; this is substitutability. Conversely, some rewards, such as food and water, are less substitutable and increases in the price of one do not necessarily engender greater consumption of the other (Green & Rachlin, 1991; Rachlin et al., 1976). In the present experiment, the exchange parameter revealed that Option B (an unsignaled and probabilistic food reward) was more substitutable for a similarly unsignaled and probabilistic food reward (Option A in the Unsignaled group).

However, Option B was less substitutable for Option A in the Signaled group, even though Option A offered the same reward commodity (food) but with an informative signal for reward. The results therefore suggest that a Pavlovian cue altered the substitutability between two otherwise equivalent operant rewards (Barrus et al., 2015) and confirm the earlier conceptualization that a signaled alternative produces a scalar increase in reward value, albeit not through a sensitivity to reward magnitude mechanism.

The results of the matching equations revealed many findings relevant to rats' dynamic decision-making. However, the matching equations offer a molar view of the potential mechanisms at work. If a scalar difference in the substitutability between options A and B accounted for group differences, the effect should also be present at the molecular trial-by-trial level as assessed by an RL model. Indeed, as suggested by the model comparison results shown in Table 2.2, the addition of group-dependent exchange parameters and sensitivity to magnitude terms (Scaled Single-Learning model) substantially improved the fits relative to the Base RL model; this result suggests that a scalar difference in reward magnitude between options and across groups is present at both molar and molecular levels of decision making. Traditionally, RL models have assumed that the value updating of each commodity is scaled independently of one another (e.g., Groman et al., 2017; Marshall & Kirkpatrick, 2017; Rutledge et al., 2009) and that choices between commodities are only relative in the sense that values between commodities are compared in the decision rule. The present finding of improved model fits through altering the RL valuation equations by way of relative magnitude scaling suggests that the independence of value assignment is incorrect. One possible reason for the traditional use of independent value scaling may stem from many RL models being implemented to model behavior in isomorphic choice procedures. However, the finding that value updating is not independent of the other available commodities is congruent with the matching assumption that preference for a reward is at least partially driven by how that reward compares to other available commodities.

One of the further benefits of molecular analyses, such as those using RL, is that they can be more modular in terms of what proposed mechanisms can be incorporated. For the present experiment, it was hypothesized that wins and losses may be scaled differentially across options and groups in addition to the scalar differences in reward

magnitude. To test the hypothesis that wins and losses scaled differentially, additional α parameters were added into the model in three steps as dependent upon Option, Outcome, and their interaction. As evidenced by the change in AIC parameters across models, the inclusion of α parameters for the interaction between Option and Outcome was substantially warranted in both groups. The results of the best fitting model suggested that the Signaled group had significantly higher β parameters and scaling of loses for Option A. The increased β values suggest the Signaled group was more "on policy" and made choices more closely in line with the model-derived reward values (Gläscher et al., 2010; Marshall & Kirkpatrick, 2017). Interpretation of the increased scaling of Option-A is not readily apparent in isolation, but the general finding that wins and losses were scaled differently has been shown previously (Estle, Green, Myerson, & Holt, 2006; Kahneman & Tversky, 1979; Marshall & Kirkpatrick, 2013) and reinforces the notion that decisions are contextually-dependent. That more group effects were not present may also have been due to the inclusion of group-dependent exchange parameters in the Scaled Quad-Learning model. Once included, the RL models for each group would then solve for the remaining parameters with a group effect already built into the model. Thus, one interpretation of the absence of group effects is that the group-dependent exchange rates largely captured the variance associated with group differences.

Although no significant differences between α parameters were apparent within groups, the AIC values for the Scaled Quad-Learning model substantially favored their inclusion and the medians shown in Figure 2.5a do suggest variability in α scaling. The dissociation between model comparisons and traditional null hypothesis significance testing (NHST) also represents a crossroad in the field as the two types of analyses potentially ask fundamentally different questions (Symonds & Moussalli, 2011; Wagenmakers & Farrell, 2004). The present experiment combined both methods and, until the RL model, both analyses were concordant. The model comparison results suggest that inclusion of α values dependent on the interaction between Option and Outcome substantially improved the model's ability to predict an individual rat's choices. Alternatively, the non-parametric tests suggest that, at a group level, there was no systematic variation in how the α estimates were ranked across rats. While the "correct" analysis is likely not objectively able to be defined, model comparison approaches are

becoming the increasingly recommended analysis approach given their ability to answer research questions in a more flexible manner (Symonds & Moussalli, 2011; Wagenmakers & Farrell, 2004). The present study may also have been under powered to detect parameter differences using nonparametric methods. In order to analyze the remaining parameters via a model comparison method, each parameter would need to be fit at the group level similar to the molar analyses. However, the inclusion of group-level parameters in RL models is complicated by parameter fits being uniquely path-dependent for each subject (Daw, 2011). The exchange rate parameters were able to be fixed because they were informed from the molar matching models while no other parameter was similarly informed. As such, the necessity of performing second level analyses on RL parameters remains a limitation of the model (Marshall & Kirkpatrick, 2017; Rutledge et al., 2009). A final consideration when interpreting the parameters from RL models is that no single parameter in isolation is fully informative of how decisions are made in the context of an RL model; that is, Figure 2.5b illustrated how the different parameters come together to determine value and showed the expected ordinal directions of Option A having greater value updating for the Signaled group relative to the Unsignaled group.

Overall, Experiment 1 suggests that rats behaved as if their decisions were value-based when choosing between two food rewards that varied as a function of both relative reward rates and magnitudes. Additionally, the presence of reward-associated cues in the Signaled group biased choice towards Option A. Although the interpretation of RL parameters is limited by both the structure of the model and the current experiment's statistical power, the molar matching analyses provided an excellent description of the data that was supported in the RL models.

Table 2.1. Experiment 1 molar models.

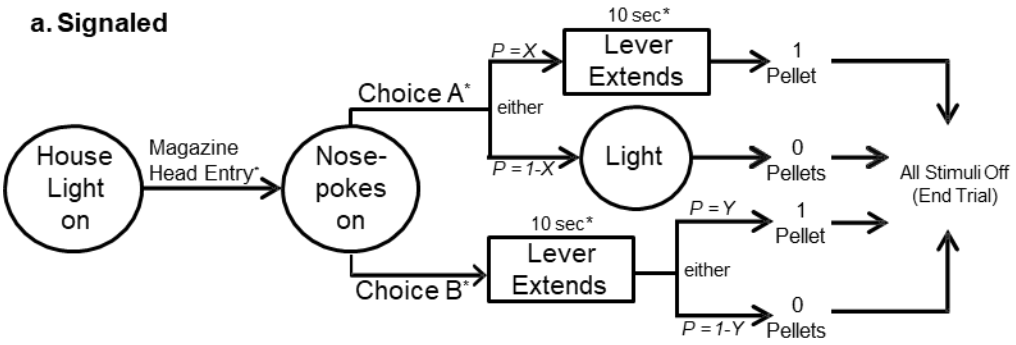
Equation	Model Name	Parameter Count	Parameters	AIC	ΔAIC
15	2-D Matching	2	S_R, S_M	-262.96	--
16	2-D Global Exchange Matching	3	S_R, S_M, Ex	-280.35	-17.39
16	**2-D Exchange Matching**	**4**	$\boldsymbol{S_R}$, $\boldsymbol{S_M}$, $\boldsymbol{Ex}$	**-315.61**	**-35.26**

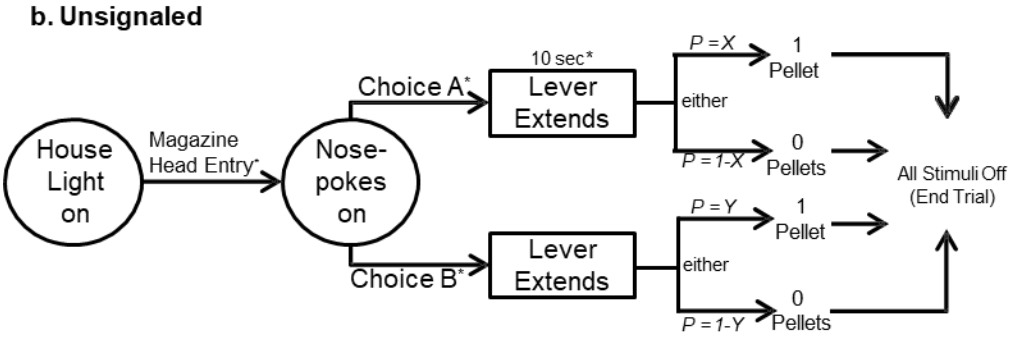
Note: Models are shown in descending order of AIC values. ΔAIC calculations are shown as the difference from the previously best fitting model. Bolded text indicates the best overall fitting model.

Table 2.2. Experiment 1 molecular models.

Model	Parameter Count	Parameters	Signaled AIC	Signaled ΔAIC	Unsignaled AIC	Unsignaled ΔAIC
Base	*3*	α,β,c	2045.45	--	2165.65	--
Scaled Single-Learning	*4*	α,β,c,S_M	1698.46	-349.99	2072.18	-93.47
Scaled Dual-Learning (Outcome)	*5*	$\alpha_{Win},\alpha_{Loss},\beta,c,S_M$	1696.59	-1.87	2064.43	-7.75
Scaled Dual-Learning (Option)	*5*	$\alpha_A,\alpha_B,\beta,c,S_M$	1651.23	-45.36	2058.09	-6.34
Scaled Quad-Learning (Option×Outcome)	*7*	$\boldsymbol{\alpha_{AWin},\alpha_{ALoss},\alpha_{BWin},\alpha_{BLoss},\beta,c,S_M}$	**1579.65**	**-71.58**	**2042.17**	**-15.92**

Note: Models are shown in descending order of overall average AIC values. ΔAIC calculations are shown as the difference from the previously best fitting model. Bolded text indicates the best overall fitting model.





c.

Probability (*P*)	X	0.66	0.22	0.11	0.11
	Y	0.11	0.11	0.22	0.66
Relative Reward Rates	A	6	2	0.50	0.17
	B	0.17	0.5	2	6

Figure 2.1. Experiment 1 general methods. (**a**) Sequence of events during initial training for the Signaled group. (**b**) Sequence of events during initial training for the Unsignaled group. The reward magnitude manipulations subsequently increased the reward magnitude for Option A to two and then three pellets. * indicates that all stimuli offset following the event. (**c**) The probabilities of events and the relative reward rates (defined as the reward probability for Option A / Option B) for Option A and B.

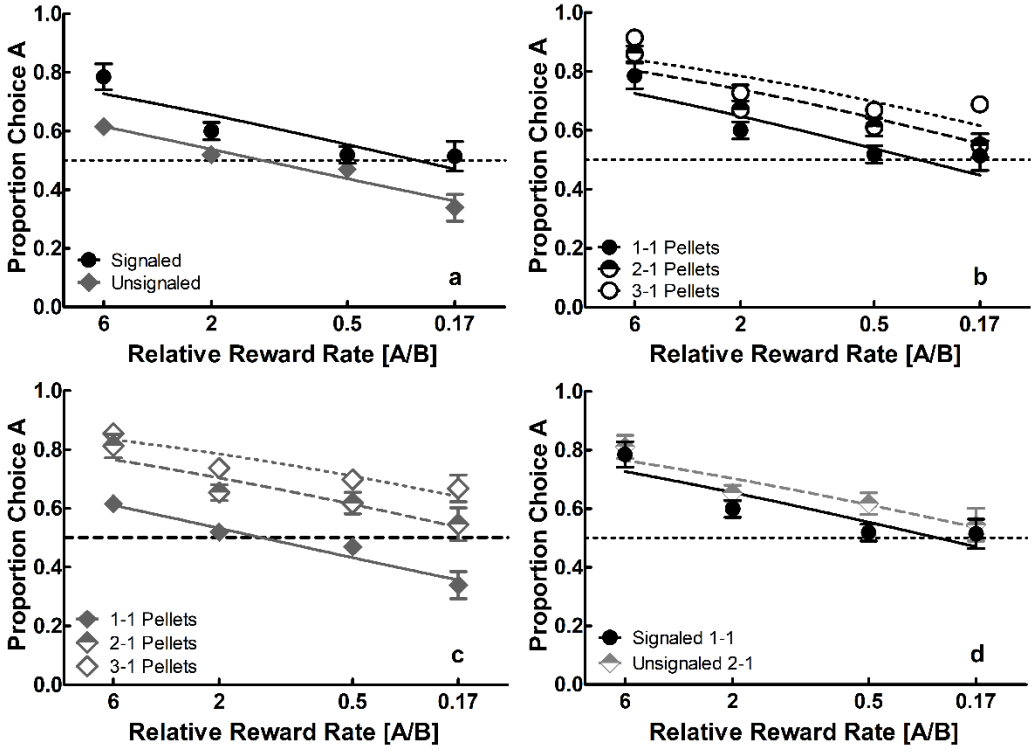


Figure 2.2. Experiment 1 molar choice data. (**a**) Mean (± SEM) proportion of choices for Option A as a function of Option A's relative reward rates for the Signaled and Unsignaled groups during initial training. Lines are the NLME-determined best fits of Equation 14. (**b**) Mean (± SEM) proportion of choices for Option A as a function of Option A's relative reward rates for the Signaled group during the magnitude manipulations. 1-1, 2-1, and 3-1 refer to the scheduled reward magnitude for Options A and B, respectively. Lines are the NLME-determined best fits of Equation 16. (**c**) Mean (± SEM) proportion of choices for Option A as a function of Option A's relative reward rates for the Unsignaled group during the magnitude manipulations. Lines are the NLME-determined best fits of Equation 16. (**d**) Mean (± SEM) proportion of choices for Option A as a function of Option A's relative reward rates for the Signaled group during the 1-1 magnitude condition and the Unsignaled group during the 2-1 magnitude condition. Lines are the NLME-determined best fits of Equations 14 and 16 for the Signaled and Unsignaled group, respectively. Note: x-axes are logarithmic for visualization.

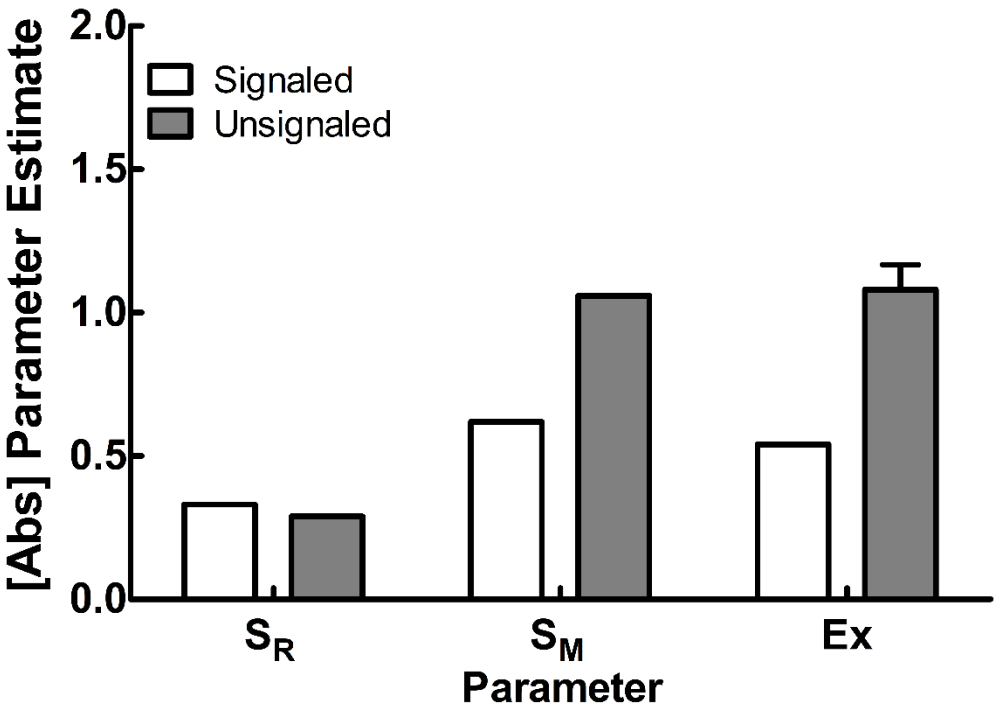


Figure 2.3. Experiment 1 molar parameter estimates. Mean (± SEM) parameter estimates from the NLME model of Equation 16 to the molar choice data during the magnitude manipulations in Experiment 1. Abs = absolute value transform. Note: error bars represent the between-subject variability of the predicted parameters. A lack of error bars reflects no differential predictions across rats for the relevant parameter.

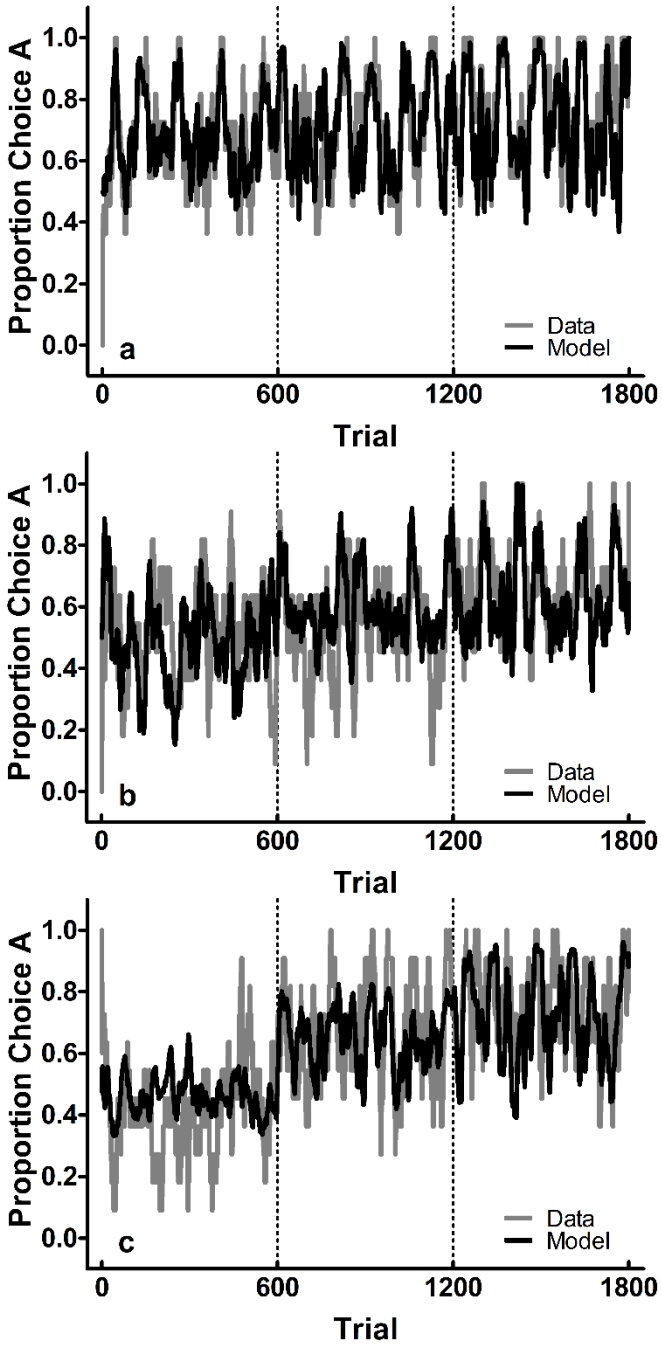


Figure 2.4. Experiment 1 RL model fits**.** Example Scaled Quad-Learning model fits to individual rat's choices of Option A. Dashed vertical lines represent reward magnitude changes for Option A in order of 1, 2, and 3 food pellets. Panels show (**a**) a Signaled rat one SD below, (**b**) a Signaled rat at, and (**c**) an Unsignaled rat one SD above the average AIC. All data were smoothed with an 11-trial moving average.

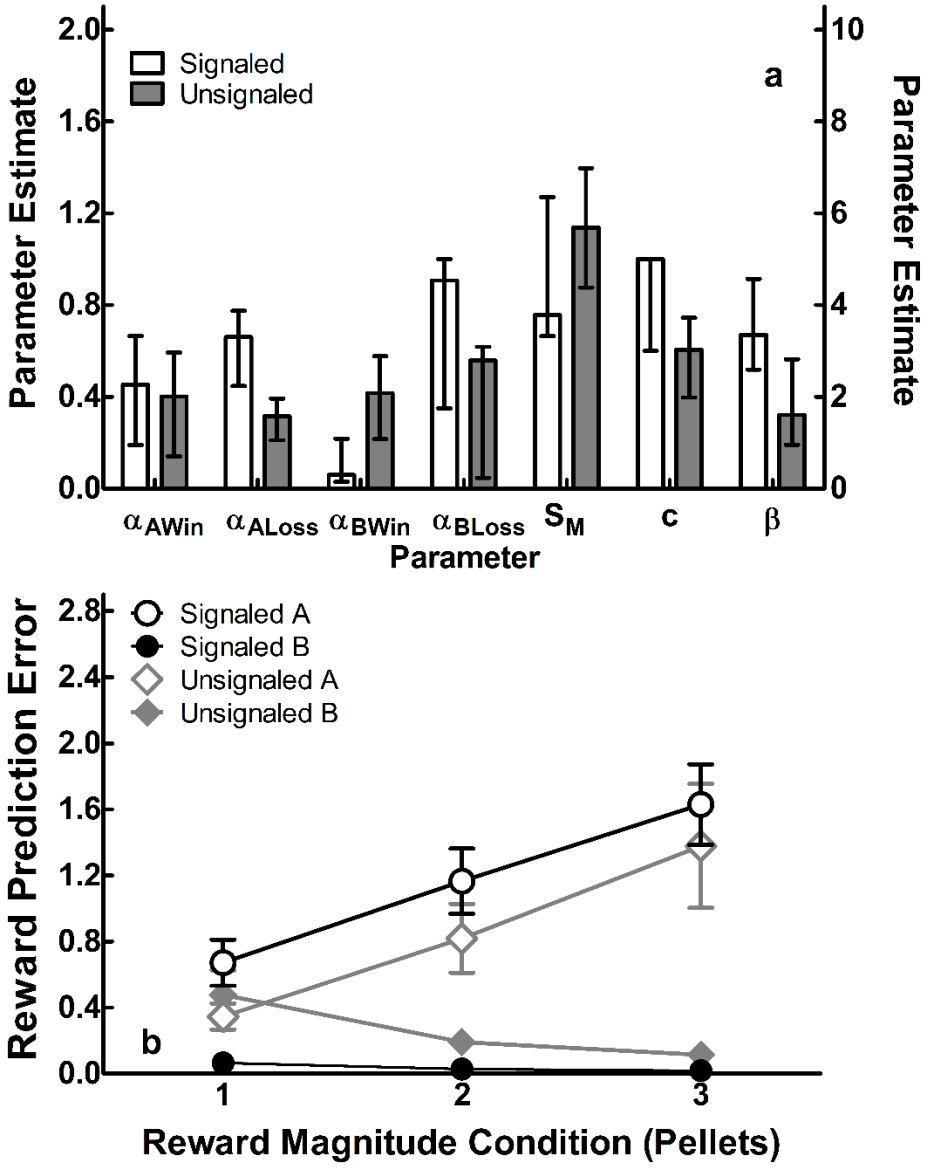


Figure 2.5. Experiment 1 molecular parameters**.** (**a**) Median parameter estimates from the Scaled Quad-Learning model to the molecular choice data in Experiment 1. Medians are shown due to the use of nonparametric test statistics, and error bars represent the interquartile range. Note: β values are scaled according to the second y-axis at the right. (**b**) Mean (± SEM) predicted reward prediction errors (RPEs) from the Scaled Quad-Learning model for Option A and B in the Signaled and Unsignaled groups as a function of Option A's reward magnitude. The displayed RPEs are from equations 18-19 when V_A and V_B equaled zero.

CHAPTER 3:
DYANMIC DECISION-MAKING FOR ALLOMORPHIC OPIOID VS. FOOD REWARDS

The results from Experiment 1 validated what to expect using the present dynamic decision-making model in drug-naïve rats responding for natural rewards. Namely, rats showed significant choice alterations between concurrently available commodities that were dependent upon the relative reward rates, relative reward magnitudes, and the presence versus absence of reward-associated cues. In the current experiment, the same isomorphic food versus food procedure was initially trained as in Experiment 1, but, after sufficient training, the food reward associated with Option A was replaced by different doses of an intravenous infusion of the opioid μ receptor agonist remifentanil (REMI). If choices for drug reward are indeed a reflection of a disorder of choice, then choices between REMI and food should show comparable effects to that of Experiment 1. However, if drugs of abuse instead transition an agent from making economic decisions to one of compulsive drug preference (e.g., Kalivas et al., 2005; Volkow et al., 2016; Wise & Koob, 2014), then repeated experience with REMI should result in inflexible choices for drug that do not vary systematically with the experimenter-manipulated variables.

Experiment 2

Methods

Subjects

Sixteen adult male Sprague-Dawley Rats (Harlan Inc; Indianapolis, IN, USA) were used in the experiment. Rats were individually housed on a 12:12 hour light:dark cycle (lights on at 7:00 a.m.) and had free access to water but were food restricted to 15 g of standard lab chow per day (Harlan Inc.) post session. All research was approved by the University of Kentucky Institutional Animal Care and Use Committee (Protocol # 2011-0885).

Apparatus

The apparatus was the same as Experiment 1 with the addition of an infusion pump (PHM-100) attached to the outside of the sound-attenuating chamber. A syringe was loaded into the pump and delivered drug through tubing inside of a metallic leash (PHM-110-SAI) that was attached to a swivel above the operant chamber. The rate of infusion delivery was 0.1 mL / 5.9 s.

Drugs

Remifentanil hydrochloride gifted from the National Institute on Drug Abuse (Bethesda, MD, USA) was mixed in sterile saline (0.9% NaCl).

Establishing operations

All establishing operations were identical to Experiment 1.

Isomorphic food versus food choice

Following establishing operations, the rats were trained on the same isomorphic dynamic decision-making procedure described in Experiment 1 and were split into Signaled and Unsignaled groups prior to training. The reward magnitudes for both Option A and B were 1 sucrose pellet and training continued for 70 sessions.

Catheter surgery

Following training on the isomorphic procedure, rats underwent surgery for implantation of a chronic, indwelling jugular catheter. Rats were anesthetized with a mixture of ketamine (Schein, Dublin, OH), xylazine (Akorn, Inc. Decatur, IL), and acepromazine (Boehringer Ingelheim, St. Joseph, MO; 75/7.5/0.75 mg/kg) delivered at 0.15 mL/100 g body weight (i.p.). The catheters were inserted into the jugular vein, extended under the skin, and exited the body via an incision on the back attached to a cannula. All animals were given 7 days to recover after surgery.

Drug self-administration training

After recovery, the rats were trained to self-administer a 3 μg/kg infusion of REMI. For both groups, Option A became associated with the REMI reward and was counterbalanced between the left and right nosepoke across animals. At the beginning of a session, the drug-associated nosepoke on the back panel illuminated and, after one response, the nosepoke's associated lever on the front panel extended into the chamber.

The lever remained extended for 10 s and the 3 μg/kg infusion began at 8.23 s into the interval and finished after 1.77 s concurrent with lever retraction. Trials were separated by a 10-s ITI. Sessions lasted for 1 hr, and training continued until all rats administered at least 10 infusions for two consecutive days.

Food versus drug training

Following initial training, sessions began by the illumination of the house-light, that required an orienting response to the magazine. Once an orienting response was made, the house light was offset and either the drug-associated Option A nosepoke or the food-associated Option B nosepoke illuminated on the back panel of the chamber. After a single response, the associated lever extended into the chamber for 10 s and an infusion of REMI or delivery of a food pellet occurred as previously described. Training continued until all rats finished all trials for two consecutive sessions.

Allomorphic drug versus food procedure

After training to self-administer both REMI and food, rats chose between a probabilistic REMI and probabilistic food reward in the same manner as described in Experiment 1. The only exception was that choice of Option A delivered a 3 μg/kg infusion of REMI rather than a sucrose pellet, such that the infusion completed concurrently with the offset of Option A's associated terminal-link stimulus. All rats were trained for a maximum of 10 days on the 3 μg/kg dose but, if a subject's average proportion of choices for Option A over five days significantly decreased as a function of Option A's relative reward rates, could be moved on to other doses after 7 days.

Reward magnitude manipulations

After training on the initial 3 μg/kg dose, the reward magnitude for Option A was changed to either 10 μg/kg or 1 μg/kg in a counterbalanced fashion. Half of the rats from each group first experienced the 1 μg/kg dose and then 10 μg/kg and half in the opposite order. Dose was manipulated by changing the pump delivery time. For the 10 μg/kg condition, the pump turned on 4.1 s into the 10 s terminal link interval while the pump turned on 9.41 s into the 10 s interval for the 1 μg/kg condition. All infusion times finished at the end of the 10 s interval concurrent with terminal link stimulus offset. Each dose was trained for a total of 10 sessions prior to moving on to the next dose.

Data Analysis

Molar analysis

Data analysis for the present Experiment closely followed Experiment 1. Choice behavior was calculated as the five-session average proportion of choices for Option A as a function of Option A's relative reward rates. The data were again analyzed using the NLME package in R and took the form of the same three derivations of the generalized matching equations (Equations 14-16) with the exception that the continuous factor of magnitude was replaced by the continuous factor of dose in µg/kg. For the initial isomorphic training, where both options offered food reward, the generalized matching equation was used (Equation 14), and model comparisons were conducted on the allomorphic drug vs. food procedure similar to Experiment 1. Namely, potential effects of Option A's reward-associated cues were quantified as altering subjects' sensitivity to changes in dose (S_M; Equation 15) or through differences in exchange rates (Equation 16). Here, the exchange rate parameter, *Ex*, scaled the reward magnitude of Option B in units of a µg/kg REMI infusion. For all analyses of drug self-administration, the data represent the average of the last five days of training for the condition.

The same sessions were also used to calculate the average interchoice intervals (ICI), or the time between choices as a function of dose. ICIs as a function of increasing doses have been shown in the literature to monotonically increase for different drugs of abuse (Beckmann et al., 2019). The present analysis sought to confirm a monotonically increasing function of ICIs with dose. ICIs were analyzed for all trials and also for trials only following REMI-Wins (i.e., Option A) as a function of dose using a linear mixed effects (LME) model in JMP with Subject as a nominal random factor, Group as a nominal fixed factor, Trial Type as a nominal fixed factor, and Dose as a continuous fixed factor. All continuous variables were mean centered.

Molecular analysis

Molecular analysis closely followed Experiment 1. However, for all models, λ_A represented the reward magnitude for Option A in REMI µg/kg units, while λ_B represented either the magnitude for Option B in food pellets in the Base RL model or the exchange rate parameter, *Ex*, scaled in REMI µg/kg units in subsequent models. Following analysis of the Base RL model, the reward prediction error term of Equations

18 and 19 were again expressed as a ratio of each option's relative reward magnitude raised to the sensitivity to magnitude parameter (S_M). On loss trials, the ratio was recoded as zero. Subsequent models were also assessed that included additional α parameters dependent on either Option (A vs. B), Outcome (win vs. loss), or the interaction between Option and Outcome. All models were again fit using the *fmincon* optimization algorithm in MATLAB using the same parameter constraints and log likelihood measures as Experiment 1. The overall best fitting model was determined according to the lowest AIC value. To quantify any potential differences between parameters, a series of Wilcoxon rank-sum tests were conducted across parameters. In the case that the best fitting model included more than one α, statistical differences across α parameters were assessed using pairwise Wilcoxon sum-rank tests with Hochberg corrections that included the relevant factors of Option, Outcome, and their interaction, depending on the best fitting model. Finally, as a visual aid to illustrate the differences in RL modeled reward magnitudes, RPEs were calculated for both options across groups.

Results

Molar analysis

Figure 3.1a shows the proportion of choices for Option A as a function of Option A's relative reward rates during the isomorphic food versus food procedure. Similar to Experiment 1, both groups tended to prefer the option with higher overall reward rates. Using dummy coding for the Group factor (due to non-convergence with effects coding), the NLME generalized matching equation again fit the data well. Results from the NLME model showed significant sensitivity to relative reward rates [$S_R = 0.26$; $F(1, 45) = 65.75$, $p < .001$] for the Unsignaled group, with significantly greater sensitivity in the Signaled group [$S_R = 0.44$; $F(1, 45) = 13.83$, $p = .001$]. The Unsignaled group also showed relatively little bias for either alternative [$b = 1.16$], while the Signaled group showed significantly more bias for Option A [$b = 0.59$; $F(1, 45) = 12.10$, $p = .001$]. Overall, the results showed that rats in both groups were making choices based on relative reward rates prior to catheter surgery, and the Signaled group again showed a bias for Option A's reward-associated cues.

For allomorphic REMI vs. food choice, Figure 3.1b-c shows the Signaled and Unsignaled group's proportion of choices for Option A as a function of Option A's

relative reward rates and REMI dose. Similar to the isomorphic procedure, allomorphic choices showed apparent control by both reward dimensions. That is, preferences shifted as both the relative reward rates and relative reward magnitudes changed. Additionally, the Signaled group showed an apparent bias for REMI reward at all doses.

Three different models were compared with the results shown in Table 3.1. Similar to Experiment 1, the 2-D Matching Model showed relatively poor fits that were improved by adding a global exchange parameter. However, allowing the exchange parameter to vary as a function of group revealed that the 2-D Exchange Matching model was substantially favored as the best fitting model. The parameter estimates (Figure 3.2) from the best fitting model also corroborated the apparent trends in the data. Both the Signaled [$S_R = 0.29$; $S_M = 1.12$] and Unsignaled groups [$S_R = 0.21$; $S_M = 1.10$] showed comparable decreases in drug preferences as the reward rates began to favor food [$F(1, 171) = 19.8$, $p < .001$] and significantly altered their choices depending on drug doses [$F(1, 171) = 147.38$, $p < .001$]. Results also showed a significant Group effect on the exchange parameter [$F(1, 171) = 11.90$, $p = .001$], indicating that Option B had greater relative worth in units of a REMI μg/kg infusion for the Unsignaled group [$Ex = 5.86$] than the Signaled group [$Ex = 3.16$].

Figure 3.3 shows the average ICIs for all trials as well as only trials following REMI-Wins. Analysis revealed that ICIs significantly increased with dose [$F(1, 14) = 16.93$, $p < .001$], ICIs following REMI-Wins were longer overall [$F(1, 14) = 52.97$, $p < .001$], ICIs following REMI-Wins increased as a function of dose more rapidly than overall ICIs [$F(1, 14) = 17.14$, $p < .001$], and that ICIs were comparable between groups.

Molecular analysis

The molar analyses supported the use of a group-dependent exchange rate parameter and suggested that scaling the valuation term in RL models as relative reward magnitude ratios may also improve overall fits. Indeed, changing RL valuation equations to a scaled ratio of relative reward in the Scaled Single-Learning model showed overwhelming evidence of improved fits (overall ΔAIC = -348.73; Table 3.2). Including additional α parameters dependent on Option and Outcome also showed substantial improvements in model fit. However, as in Experiment 1, adding α values dependent on

the Option × Outcome interaction in the Scaled Quad-Learning model was the best fitting model.

Figure 3.4 shows three example model fits that approximate the average AIC (± one standard deviation). Overall, the RL model successfully parameterized the subjects' trial-by-trial choices [ΔAIC -640.60 from a chance model] across all manipulated variables despite its foundation in conditioning principles for non-drug commodities. Parameters from the Scaled Quad-Learning model are shown in Figure 3.5a. There were no significant group effects on any parameter, but α values for wins were significantly higher than losses across both options for the Signaled [$Z = 2.52$, $p = .012$] and Unsignaled groups [$Z = 2.52$, $p = .012$]. Finally, Figure 3.5b shows the calculated RPEs for win trials when the current value for both options equaled 0. As in Experiment 1, the calculated RPEs showed the expected ordinal effects of the Signaled group showing greater value updating for Option A than the Unsignaled group that showed greater value updating for Option B.

Discussion

The purpose of Experiment 2 was to assess the ability of rats to make value-based decisions when choosing between probabilistic drug and nondrug rewards. To that end, the results strongly support that rats in the present experiment were making decisions between REMI and food as if they were value-based. Rats in both groups showed significant sensitivity to relative reward rates as evidenced by decreasing choices for REMI when the reward rates favored food (i.e., Option B). Additionally, as the dose of REMI increased, so did choices for REMI. Finally, the Signaled group's reward-associated cues produced a significant bias for REMI across all reward rate and magnitude conditions but otherwise did not alter the effects of changes on those dimensions to allomorphic decision-making.

Prior to the allomorphic procedure, it was important to ensure that rats, during the isomorphic procedure, showed sensitivity to relative reward rate changes similar to Experiment 1. Without replicating the effects of Experiment 1, any potential deficits seen in allomorphic drug-associated decision making could be attributed to an overall lack of learning. However, rats in the current experiment's isomorphic procedure showed significant sensitivities to relative reward rates. Interestingly, the present experiment also

showed that the Signaled group had heightened sensitivities to relative reward rates compared to the Unsignaled group. Although unexpected, the increased S_R values may be interpreted in a manner similar to the group differences in sensitivity to reward magnitude from Experiment 1. That is, sensitivities to a given dimension are likely to be contextually dependent such that the two groups, through the differential use of reward-associated cues, made decisions in different contexts. The increased S_R values may also be interpreted from the perspective of differential outcomes research (Urcuioli, 2005), that has shown improved discrimination learning when two choices lead to either different stimuli or rewards.

After sufficient training on the isomorphic condition, the rats were moved on to the allomorphic drug versus food condition to assess if they could make value-based allomorphic decisions. Indeed, contrary to compulsion hypotheses of addiction (e.g., Kalivas et al., 2005; Wise & Koob, 2014), replacing Option A's associated food reward with varying REMI doses revealed that choices for REMI were sensitive to both the relative reward rates and magnitudes of the concurrently available options. That rats were sensitive to reward magnitude (i.e., dose) corroborates previous research using REMI and other opioid μ receptor agonists (Freeman & Woolverton, 2011; Koffarnus & Woods, 2008; Negus, 2006; Spiga, Maxwell, Meisch, & Grabowski, 2005), and the dose-dependent increase in the average ICIs parallels previous procedures assessing cocaine choice (Beckmann et al., 2019). However, the present study is believed to be the first to show that choices of an opioid μ receptor agonist are sensitive to relative reward rates, thus corroborating findings with cocaine choice (Woolverton & Rowlett, 1998). Experiment 2 also showed that relative reward rates and magnitude produce independent effects on REMI choices in line with the predictions made by the matching literature (e.g., Kyonka, 2008) outlined in Experiment 1. Namely, a single sensitivity parameter for relative reward rates and magnitudes sufficiently described the data. Further, the average parameter estimates for sensitivity to relative reward rates were similar between the current experiemnt (mean $S_R = 0.25$) and Experiment 1 (mean $S_R = 0.30$), suggesting that the dimensional control exerted by the relative reward rates was comparable between isomorphic food-food and allomorphic drug-food decision-making conditions.

Also similar to Experiment 1, the Unsignaled group made decisions as if Option B's reward magnitude was larger (*Ex* = 5.86) than the Signaled group (*Ex* = 3.16) despite choice of Option B resulting in an identical reward magnitude (i.e., one food pellet). Scaled in units of a REMI μg/kg infusion, the group effect on the exchange rate parameter suggests a scalar difference in how substitutable (Green & Freed, 1993) an unsignaled food pellet is relative to an unsignaled REMI infusion (for the Unsignaled group) or a signaled REMI infusion (for the Signaled group). Thus, whether choosing between identical rewards (Experiment 1) or qualitatively different rewards (current experiment), the reward-associated cues produced a similar effect on how substitutable Options A and B are for one another. The present experiment is also believed the first to show that reward-associated cues as used here can bias choice for drug reward and corroborates previous research using food rewards (Zentall, 2016) as well as research showing that conditioned stimuli can alter the value of drug rewards (e.g., Beckmann et al., 2019; Tunstall & Kearns, 2014).

Given that the molar data were suggestive of a scalar difference in reward magnitude between REMI and food across the two groups, the molecular RL analysis sought to confirm the finding at a trial-by-trial level. As in Experiment 1, the RL model comparisons supported scaling the valuation equations as a ratio of relative reward magnitudes that included a group-dependent exchange term and sensitivity to reward magnitude parameter. As discussed in Experiment 1, the implications of altering the RL valuation equations from independent to relative reward magnitude represents an important difference from traditional RL models that only recognize the relative nature of choice preferences in the decision equation. The current experiment extends upon Experiment 1 by providing a replication of the effect in allomorphic choice conditions. Additionally, for the same reasons as Experiment 1, the lack of any group effects on RL parameters may not be without cause. That is, inclusion of group-dependent exchange terms in the Scaled Quad-Learning model effectively builds a group effect into the model around which the remaining parameters may optimize. Therefore, once the scalar differences in reward magnitude of the two options are accounted for, other decision parameters may have become somewhat consistent.

The RL model also extends the results of the molar matching models through its ability to scale wins and losses with the α parameters. As in Experiment 1, the inclusion of additional α parameters was beneficial for model fits according to AIC differences, and the non-parametric tests showed that wins were weighted significantly more than losses for both alternatives in both groups. That wins and losses were differentially scaled suggests there was variance in how rats in the present experiment made decisions that went beyond what was captured by the reward magnitude differences across options (Kahneman & Tversky, 1979; Marshall & Kirkpatrick, 2017). Additionally, it is believed that the present experiment is the first to apply RL models to drug choices in any species and among a minority that do so for rodent choice data (e.g., Groman et al., 2017; Marshall & Kirkpatrick, 2017). The finding that an RL model can successfully parameterize drug choices is a significant finding, as RL models imply that an agent is calculating expected payoffs and adjusting future choices accordingly. Such value-based decision-making is certainly not something one would expect from a compulsive drug user.

In terms of the current result providing a satisfactory test of compulsive drug use hypotheses, two potential limitations were the choice of drug and whether or not the current rats represented "addicted" individuals. Indeed, although REMI is an opioid μ receptor agonist similar to other abused opioids (Glass et al., 1993; SAMHSA, 2017; Trescot, Datta, Lee, & Hansen, 2008), its relatively short half-life (Crespo et al., 2005; Glass et al., 1993; Haidar et al., 1997) is thought to contribute to a lower abuse liability (Baylon, Kaplan, Somer, Busto, & Sellers, 2000). However, as clearly evidenced by rats' choices in the present experiment, REMI has reinforcing efficacy. Furthermore, at least one abuse case of REMI exists in humans (Levine & Bryson, 2010). Future research could therefore expand upon the current experiment through use of other opioid μ agonists.

As to whether the present rats represent "addicted" individuals, the allostatic theory of drug addiction (e.g., Wise & Koob, 2014) suggests that choices for a drug are initially driven by the positive reward value of such drugs of abuse. After repeated administrations, tolerance to the pharmacological effects develop that both diminish the drug's positive rewarding value and alters the individual's biological homeostatic set

point. During this later stage of addiction (i.e., tolerance), drug use is driven to maintain the new homeostasis point reliant on the drug's presence and to avoid the associated withdrawal consequent upon failure to maintain drug use. Ultimately, the result is that the user transitions from voluntary to compulsive use where drug taking occurs independently of the drug being hedonically pleasing. In rats, the transition to compulsive use is ostensibly modeled using the previously described escalation procedures (Ahmed & Koob, 1998; Lenoir et al., 2013), where rats who are given extended access to a drug show increased levels of drug consumption when offered in isolation.

That the present rats did not undergo an escalation of drug intake procedure prior to making allomorphic choices may suggest that the current rats simply do not meet the criteria for addiction. However, the suggestion that individuals are only compulsive when they are addicted is a cyclical mode of thinking that precludes hypothesis testing of addiction theories. Additionally, evidence that escalation of drug intake reflects dysregulated and compulsive drug use remains speculative. Choice preferences for stimulants after evidence of escalation remains unchanged (Caprioli et al., 2015; Lenoir et al., 2007), and there is only one report showing that preferences for heroin increases after escalation (Lenoir et al., 2013). Furthermore, although escalation commonly shows that drug intake during a short-access period (i.e., 1 hr) is regulated (i.e., does not escalate), rats who are first acclimated to a 10-min self-administration period have shown escalation during a subsequent 1-hr self-administration session (Beckmann, Gipson, Marusich, & Bardo, 2012). Rats have also shown evidence of escalating food consumption (Goeders, Murnane, Banks, & Fantegrossi, 2009; Venniro, Zhang, Shaham, & Caprioli, 2017), and training rats on an escalation procedure prior to an allomorphic decision-making procedure produces a strongly biased reward history that is often not acknowledged (Beckmann et al., 2019). Additionally, although there is mixed evidence indicating that animals experiencing opioid withdrawal (i.e., a marker of biological drug dependence) show increased drug preferences (Lenoir et al., 2013; Negus, 2006; Negus & Rice, 2009), rats who have been acutely deprived of food will also show increased food preference and hoarding (Beckmann et al., 2019; Morgan, Stellar, & Johnson, 1943). Thus, the ostensible behavioral markers of dysregulated drug use may instead

reflect general learning and behavioral mechanisms that do not discredit rats' value-based decision-making for a REMI food reward in the present experiment.

To summarize, the current experiment revealed several findings relevant to our understanding of drug abuse. First, choices for REMI were shown to be dependent upon both the relative reward rates and magnitudes of the concurrently available commodities. Second, the presence of conditioned stimuli associated with REMI reward significantly biased decision making towards REMI choices in a manner similar to stimuli being paired with food reward. Third, the allomorphic choice dynamics were able to be captured at both the molar level with matching models and the molecular level with RL models. Finally, the present experiment (1) suggests that rats are indeed capable of making value-based decisions for drug commodities and (2) establishes a basic procedure from which to test further decision-making hypotheses in a more dynamic procedure than is typical in substance abuse research.

Table 3.1. Experiment 2 molar models.

Equation	Model Name	Parameter Count	Parameters	AIC	ΔAIC
15	2-D Matching	2	S_R, S_M	-148.71	
16	2-D Global Exchange Matching	3	S_R, S_M, Ex	-231.40	-82.69
16	**2-D Exchange Matching**	**4**	$\boldsymbol{S_R}$**,** $\boldsymbol{S_M}$**,** $\boldsymbol{Ex}$	**-321.86**	**-90.46**

Note: Models are shown in descending order of AIC values. ΔAIC calculations are shown as the difference from the previously best fitting model. Bolded text indicates the best overall fitting model.

Table 3.2. Experiment 2 molecular models.

Model	Parameter Count	Parameters	Signaled AIC	Signaled ΔAIC*	Unsignaled AIC	Unsignaled ΔAIC*
Base	*3*	α,β,c	2246.41		2372.23	
Scaled Single-Learning	*4*	α,β,c,S_M	1957.93	-288.48	1962.25	-408.98
Scaled Dual-Learning (Outcome)	*5*	$\alpha_{Win},\alpha_{Loss},\beta,c,S_M$	1936.68	-21.25	1923.14	-39.11
Scaled Dual-Learning (Option)	*5*	$\alpha_A,\alpha_B,\beta,c,S_M$	1932.58	-4.1	1924.25	1.11
Scaled Quad-Learning (Option×Outcome)	*7*	$\boldsymbol{\alpha_{AWin},\alpha_{ALoss},\alpha_{BWin},\alpha_{BLoss},\beta,c,S_M}$	**1867.79**	**-64.79**	**1841.66**	**-82.59**

Note: Models are shown in descending order of overall averaged AIC values. ΔAIC calculations are shown as the difference from the previously best fitting model. Bolded text indicates the best overall fitting model.

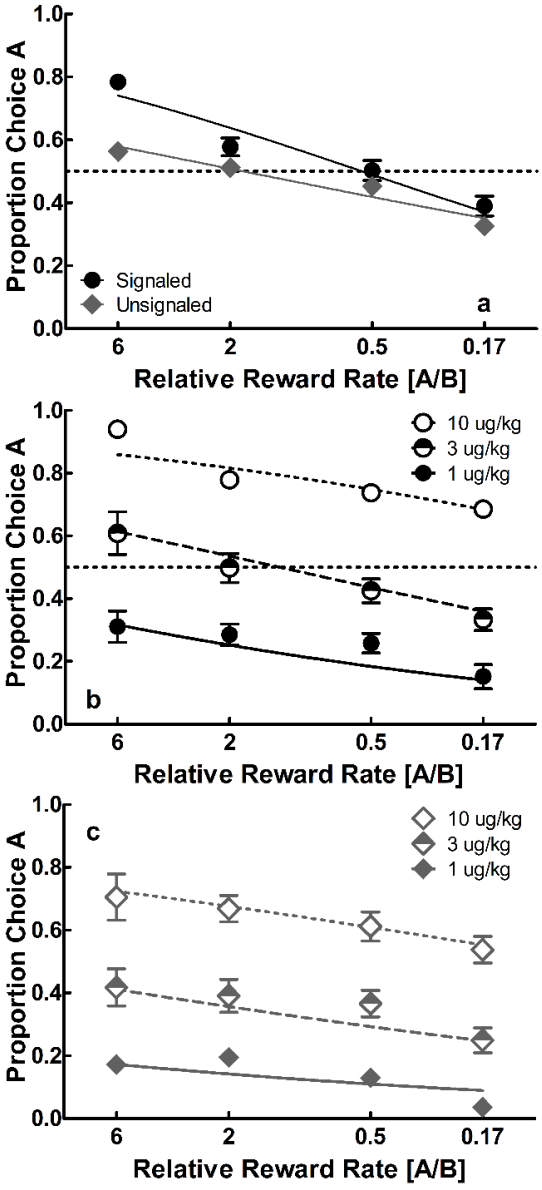


Figure 3.1. Experiment 2 molar choice data. (**a**) Mean (± SEM) proportion of choices for Option A as a function of Option A's relative reward rates for the Signaled and Unsignaled groups during the isomorphic food versus food procedure. Lines are the NLME-determined best fit of Equation 14. (**b-c**) Mean (± SEM) proportion of choices for Option A as a function of Option A's relative reward rates and REMI dose for the Signaled (**b**) and Unsignaled (**c**) group during the allomorphic drug vs. food procedure. Lines are the NLME-determined best fit of Equation 16. Note: x-axes are logarithmic for visualization.

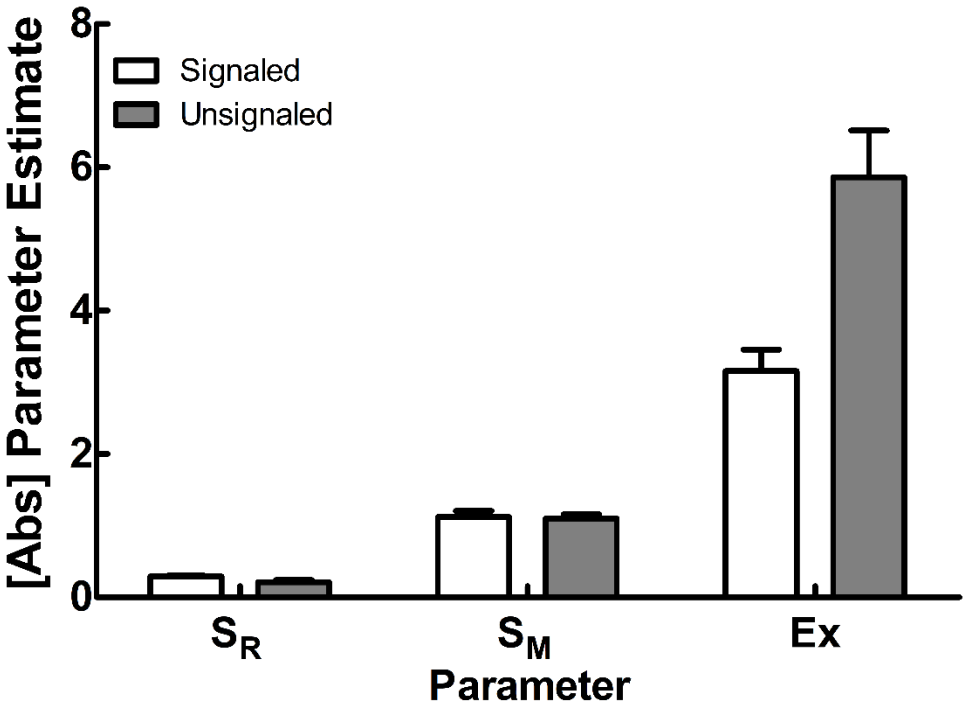


Figure 3.2. Experiment 2 molar parameters**.** Mean (± SEM) parameter estimates from the NLME-determined best fit of Equation 16 to the molar choice data during the allomorphic REMI vs. food procedure in Experiment 2. Abs = absolute value transform. Note: error bars represent the between-subject variability of the predicted parameters.

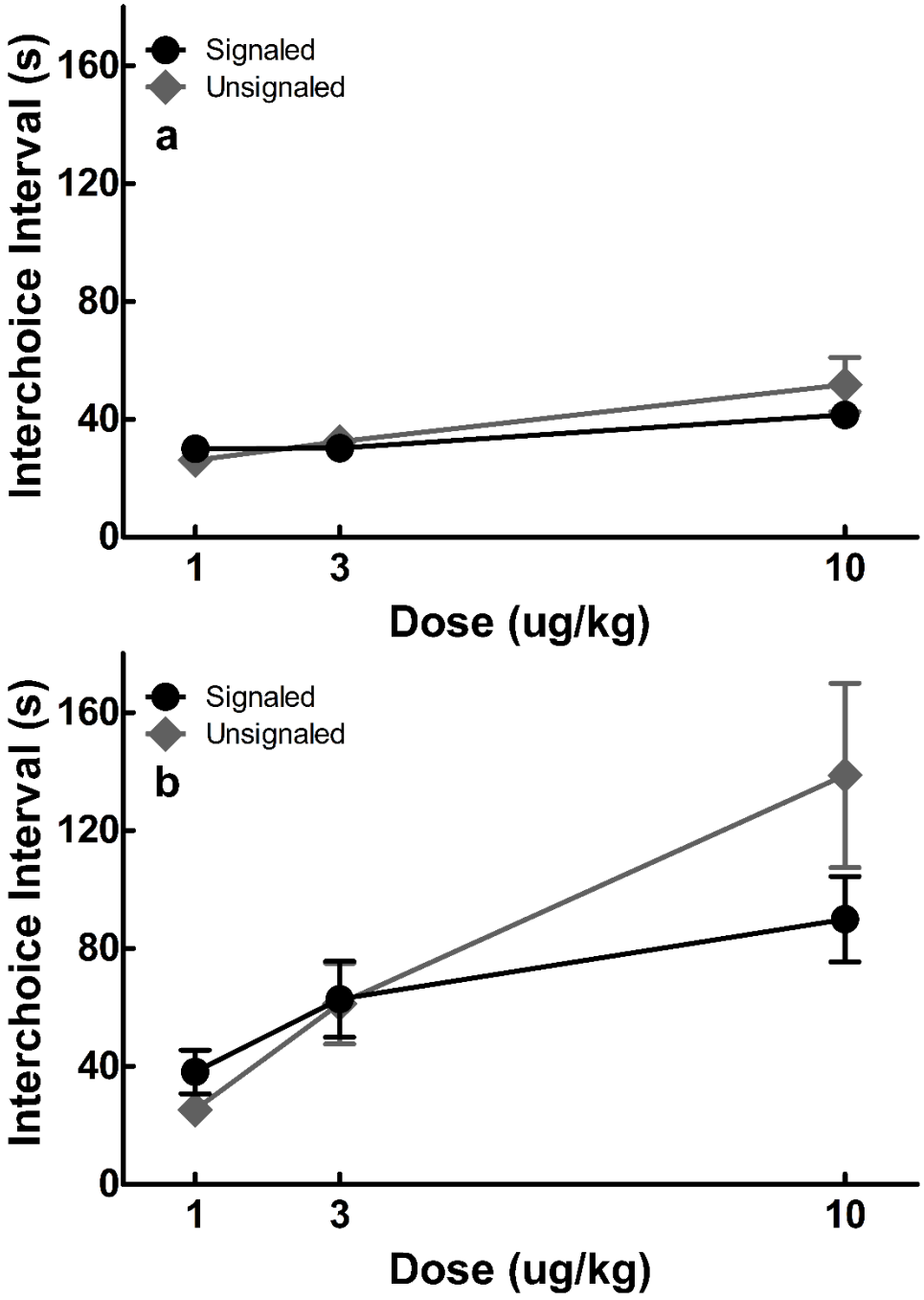


Figure 3.3. Experiment 2 interchoice intervals**.** (**a**) Mean (± SEM) interchoice intervals for all trials in the Signaled and Unsignaled group as a function of dose. (**b**) Mean (± SEM) interchoice intervals for trials following a rewarded choice for Option A (i.e., a REMI-Win) in the Signaled and Unsignaled groups as a function of dose.

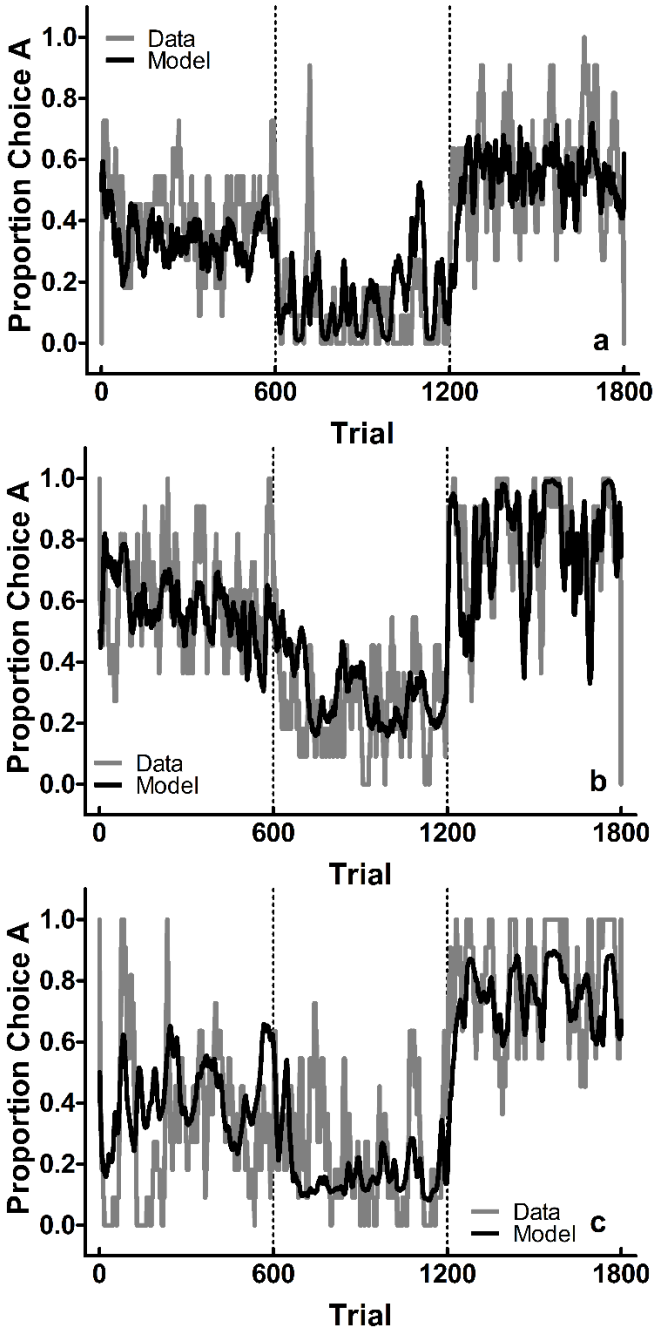


Figure 3.4. Experiment 2 RL model fits. Example fits from the Scaled Quad-Learning model to molecular choice data of Option A. Dashed vertical lines represent reward magnitude changes for Option A in order of 3, 1, and 10 µg/kg doses. Panels show (**a**) an Unsignaled rat one SD below, (**b**) a Signaled rat at, and (**c**) a Signaled rat one SD above average the AIC. All data were smoothed with an 11-trial moving average.

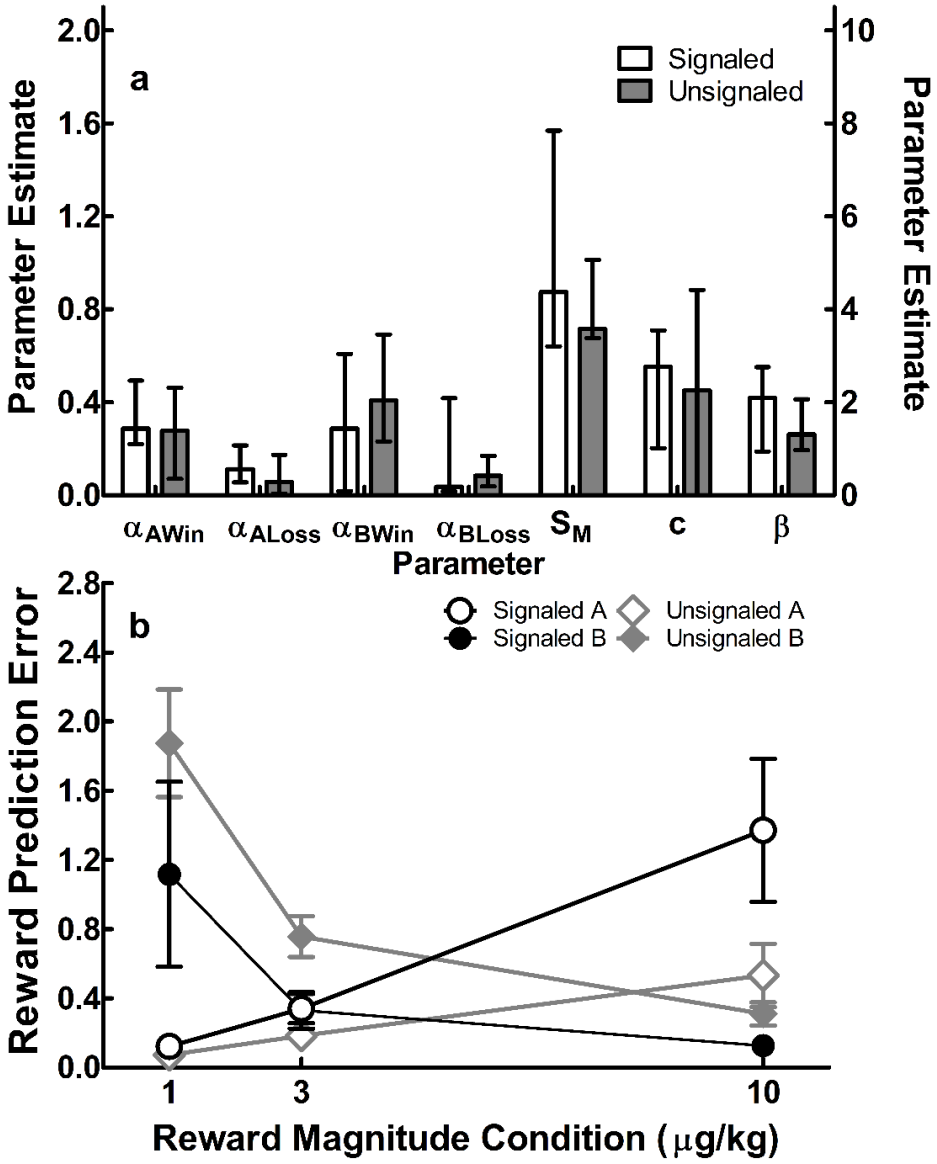


Figure 3.5. Experiment 2 molecular parameters. (**a**) Median parameter estimates from the Scaled Quad-Learning model fit to the molecular choice data in Experiment 2. Medians are shown due to the use of nonparametric test statistics, and error bars represent the interquartile range. Note: β values are scaled according to the right y-axis. (**b**) Mean (± SEM) predicted reward prediction errors (RPEs) from the Scaled Quad-Learning model for Options A and B in the Signaled and Unsignaled group as a function of Option A's reward magnitude. The displayed RPEs are from equations 18-19 when V_A and V_B equaled zero.

CHAPTER 4
NEUROECONOMIC ASSESSMENT OF REWARD VALUE DURING ALLOMORPHIC OPIOID VS. FOOD DECISION-MAKING

The results of Experiment 2 suggest that rats made allomorphic drug versus food choices according to value-based decision-making mechanisms. As such, a logical extension of Experiment 2 is to ask whether such allomorphic decisions are mediated by the scaling of brain activity to the different dimensions modulating the propensity for drug choices (e.g., Glimcher et al., 2005). As such, oxygen-sensitive electrodes were implanted into the nucleus accumbens (NAc) of rats engaging in an allomorphic decision-making procedure as a proxy for fMRI BOLD signals (Francois et al., 2012; Lowry et al., 2010). However, in order to maximize the ability to detect NAc O_2 scaling with relative reward rates, the signaling condition of Experiment 3 was changed to include distinct win- and loss-associated stimuli for both options; this was done due to the finding of improved sensitivity to relative reward rates in Experiment 2 for the Signaled group during the isomorphic procedure. If NAc activity encodes a putative reward value signal, then activity should scale with the modulators of drug choice such as relative reward rates and magnitudes.

Experiment 3

Methods

Subjects

Six adult male Sprague-Dawley Rats (Harlan Inc; Indianapolis, IN, USA) were used in the experiment. Rats were individually housed on a 12:12 hour light:dark cycle (lights on at 7:00 a.m.) and had free access to water but were food restricted to 15 g of standard lab chow per day (Harlan Inc.) post session. All research was approved by the University of Kentucky Institutional Animal Care and Use Committee (Protocol # 2011-0885).

Apparatus

The apparatus was the same as in Experiment 2.

Drugs

Remifentanil hydrochloride gifted from the National Institute on Drug Abuse (Bethesda, MD, USA) was mixed in sterile saline (0.9% NaCl).

Oxygen biosensor

Microelectrode arrays (MEAs, S2 configuration; CenMeT, Univeristy of Kentucky) were built as previously described (Rutherford, Pomerleau, Huettl, Strömberg, & Gerhardt, 2007) and consisted of four platinum recording sites (15 µm × 333 µm) arranged in dual pairs. Once built, the MEAs were calibrated in vitro at 100 Hz using the FAST16 mkIII electrochemical recording system (Fast Analytical Sensing Technology, Quanteon, LLC, Nicholasville, KY). In vitro calibration was adapted from Ledo et al. (2017). Briefly, a 20 mL solution of 0.05 M, pH 7.4, phosphate-buffered saline was purged of O_2 with nitrogen at 37°C. Three additions of 4.95 µM oxygen were then used to generate a calibration curve to determine MEA sensitivity (slope, nA/µM), limit of detection (in µM, signal-to-noise = 3), and linearity ($R^2 \geq 0.9$).

Establishing operations

All establishing operations were the same as Experiment 1.

Isomorphic food versus food choice

Rats were trained on the same isomorphic decision-making procedure as Experiment 1 with the exception that all rats were in a Double Signaled condition. Illustrated in Figure 4.1, Option A was equivalent to Option A for the Signaled groups in Experiments 1 and 2. However, here, choosing Option B produced a lever stimulus when a reward followed and a jewel light stimulus directly above the lever when a loss occurred for 10 s. Thus, there was no differential signaling across Option A and B. In order to reach a stable baseline, a training criterion was set such that subjects had to have S_R values greater than or equal to -0.25 to show comparable values with Experiments 1 and 2. However, equipment for subsequent training was not always available once training criterion was reached and resulted in less systematic training times. The average training time contingent on both S_R values and equipment availability was 54 sessions (SEM = 11.29).

Catheter surgery

Once the training criterion had been met, rats underwent surgery for implantation of the jugular catheter as described in Experiment 2.

Drug self-administration training

Drug self-administration training was as described in Experiment 2 with the exception that all rats were only trained on the 10 μg/kg REMI dose. Following choice of the drug-associated Option A, the infusion began 4.1 s into the 10-s terminal link interval and finished concurrent with terminal link offset.

Food versus drug training

Following completion of drug self-administration training, all rats completed food versus drug training as described in Experiment 2.

Preliminary allomorphic decision-making procedure

After food versus drug training, all rats completed approximately five days of training on the allomorphic decision-making procedure described in Experiment 2. Experiment 3 also included forced-choice trials, which were identical to free-choice trials except that, after orienting to the house-light, only one of the two nosepokes were available. Forced-choice trials were implemented in order to maintain adequate sampling of both choice options across all conditions of training and because previous research has shown dissociable effects between forced- and free-choice trials (e.g., Sugam, Day, Wightman, & Carelli, 2012). There were 12 forced choice trials per block (6 per option) for a total of 48, which reduced the total free-choice trial count (when both choice options were available) to 72 total trials.

Electrode implantation

Following completion of preliminary allomorphic decision-making training, rats underwent a second surgery for the implantation of the O_2-sensitive MEA. Prior to surgery, all rats were given a 10 mg/kg dose of carpofen (Rimadyl, Pfizer, NYC) and 1 mL of 0.9% sterile saline subcutaneously. Rats were then anesthetized using 4% isoflurane (Isothesia, Henry Schein, Melville, NY), their head was shaved, and the rat was placed in a stereotaxic frame (David Kopf Instruments, Tujunga, CA) on top of a circulating water bath to maintain body temperature at 37°C. Rats were maintained at an

isoflurane level of 1-3%, artificial tears (Henry Shcein, Melville, NY) were applied to the rats' eyes, and the shaved heads were disinfected with Hibiclens scrub (Mölnlycke Health Care, Norcross, GA) and 70% ethanol. An incision was made on the rat's head to expose the skull, followed by a craniotomy exposing the right hemisphere of the NAc (AP: +1.8-2.04; ML: ± 1.5; DV: -7.0 relative to bregma; Paxinos & Watson, 2006). Three screws (Amazon supply, part No. B00FN0K02) were then screwed into the skull and a small burr hole was made contralateral to the site of the craniotomy for implantation of the Ag/AgCl reference electrode. The MEA was then implanted into the NAc, the Ag/AgCl reference electrode into the burr hole, and the MEA was set in place using dental acrylic (Ortho Jet Powder and Jet Acrylic Liquid, Lang Dental Manufacturing Co., Wheeling IL). Rats were allowed to recover for three days prior to resuming behavioral training and were given repeated carpofen (10 mg/kg s.c.) daily.

Electrochemical recordings

Oxygen measurements were performed with the FAST-16 mkIII recording system using a low noise 4-channel Rat Hat amplifier system (Quanteon, LLC, Nicholasville, KY) connected through a low-noise commutator (Plastics One, Inc., Roanoke, VA) at 100 Hz. During behavioral training, the MEA was maintained at -0.6 V versus the Ag/AgCl reference electrode (Ledo et al., 2017).

Reestablishing operations

Following electrode implantation, rats again underwent drug self-administration training followed by drug versus food training as described above.

Allomorphic decision-making procedure

After reestablishing procedures, rats returned to the allomorphic decision-making procedure as previously described. All rats were trained for a minimum of 10 sessions per REMI dose, first with Option A providing 10 μg/kg of REMI, then 3 μg/kg, and finally 1 μg/kg. For the first five days of each dose, voltage was not applied to the electrode and no electrochemical recordings were collected. For the remainder of training at each component, voltage was applied to the MEA and recordings collected. A minimum of five recording sessions were collected for each animal but sessions could be excluded due to problems with the delivery of REMI, poor quality of the electrical

recordings, or failure to complete the session. In the case of excluded sessions, further training was necessitated. After all data were collected, animals were euthanized, the brains were extracted and flash frozen, and 40 μm brain slices were prepared using a cryostat. The slices were stained using Cresyl Violet (Sigma-Aldrich, St. Louis, MO) and visualized to confirm MEA placements into the NAc (Figure 4.15).

Data Analysis

Molar analysis

Molar data analysis closely followed Experiment 2. Choice data were expressed as the proportion of choices for Option A as a function of Option A's relative reward rates. To quantify the initial isomorphic food versus food data, a NLME model taking the form of Equation 14 was used in the NLME package in R. To quantify the allomorphic drug versus food data, a NLME model taking the form of the Global Exchange 2-D Matching Equation (Equation 16) was used. No model comparisons were conducted due to the absence of groups. All molar choice data reflect the average of five sessions, and no forced choice trials were included in molar choice data analysis. The same sessions used for the molar choice data analysis were used to analyze ICIs in the same manner as Experiment 2, except there was no Group factor in the LME model.

Molecular analysis

Molecular analysis was the same as Experiment 2 except no group comparisons were conducted on the parameter estimations. Additionally, on forced choice trials, the value of both options was updated but no choice was predicted. Instead, the last predicted choice was retained for the forced choice trial and did not count towards log-likelihood estimations (Marshall & Kirkpatrick, 2017). Analysis of RL parameters was only conducted if the best fitting model included more than one α parameter; in this case, a series of pairwise Wilcoxon signed-rank tests were run with Hochberg corrections that included the relevant factors of Option, Outcome, and their interaction. Finally, RPEs for both options were calculated as a visual aid to illustrate the differences in model value updating.

Oxygen analysis

Changes in oxygen dynamics were analyzed for each trial by first creating a 40-s trial window around the time of a choice. The trial window was then segmented into three

separate intervals: baseline, cue, and reward. Baseline intervals were defined as 4 s prior to a choice until 1 s prior to a choice (3 s total). Cue intervals were defined from the time a choice was made until the end of the 10-s terminal link stimulus (10 s total). Reward intervals were defined as 1.5 s prior to the offset of the terminal link stimulus (in order to capture large, convoluted signals between cue and reward intervals) until 5 s past the ITI (27.5 s total). However, in the case that the next trial was initiated prior to the end of the reward interval, the reward interval was truncated to 3 s before the initiation of the next trial (i.e., orienting response). All trial windows were smoothed with a 0.25 s moving average.

Trial windows were then analyzed using the *findpeaks* function in MATLAB, which identifies local maxima (i.e., peaks) in timeseries data separately for the cue and reward intervals. All local maxima were required to have a minimum duration of 0.75 s, and the largest returned value was recorded as the peak for a given trial's cue or reward interval. Furthermore, given that RPEs can be both positive and negative (e.g., Schultz, 2010), each trial window was analyzed again to search for local minimums (i.e., troughs) by multiplying the trial window by -1. However, early in the analysis it became apparent that REMI-Wins in the 10 μg/kg condition produced a very large O_2 response that extended well beyond the reward interval previously described. As such, reward intervals for REMI-Wins in the 10 μg/kg condition were set to 190 s from the time of choice. Additionally, to aid the *findpeaks* function in capturing the response to REMI 10 μg/kg infusions, the trial window was filtered with an elliptic lowpass filter using the *ellip* function in MATLAB. The lowpass filter stopband was set to 0.09 Hz and was chosen through experimentation with exemplar trials due to its minimal alteration of the overall signal. In the case that the next trial was initiated prior to the end of the 190-s reward interval, the reward interval was truncated to 3 s prior to the trial's initiation.

The dependent variables collected from each cue and reward interval of a given trial were the largest-amplitude peak and trough (i.e., maximum and minimum, respectively), expressed as a percent change from the mean of the baseline interval, the width of the largest-amplitude peak and trough, and the proportion of trials in which a peak or trough was found. The dependent variables were then entered into a linear mixed-effects (LME) model in MATLAB with Subject as a nominal random factor, Dose as a

continuous fixed factor, Block (i.e., relative reward rates) as a continuous fixed factor, Option as a nominal fixed factor, Trial Type (free vs. forced) as a nominal fixed factor, and Outcome as a nominal fixed factor. All nominal variables were effects coded and continuous variables mean centered. The primary interest of the current study was how any changes in the O_2 signal scaled with either relative reward rates (i.e., Block) or dose manipulations of REMI. In the case of significant effects coupled with significant interactions, the highest order interaction was interpreted. For post hoc tests of interactions, coefficient tests were run using the *coeftest* function in MATLAB. Finally, in the case of apparent quadratic trends, quadratic trend analyses were conducted in JMP. To differentiate the two sets of analyses, reporting the post hoc testing in MATLAB herein involves the term *slope* for the testing of linear simple slopes, while those done in JMP herein involves the term *dose* and $dose^2$.

Important limitations with the MEA data collected warrant discussion. For instance, two out of six animals displayed repeated saturation of the MEA signal that precluded collection during select time periods, one animal displayed behavioral stereotypy during win trials that produced electrical disturbances, and one MEA changed its resting baseline values within session. For these reasons, the entire electrical signal for each session was first visualized prior to analysis. Upon detecting nonfaradaic current, any trials where the aforementioned electrical disturbances existed were manually excluded. The average percent of trials excluded across subjects was 13.21% (SEM = 5.12%).

Results

Molar analysis

Figure 4.2a shows the proportion of choices for Option A as a function of Option A's relative reward rates during the isomorphic food versus food procedure. Similar to previous experiments, the rats tended to favor the options with greater relative reward rates. Analysis with the general matching NLME model also confirmed the rats' choices were sensitive to the changing relative reward rates [$S_R = 0.45$, $F(1, 17) = 106.08$, $p < .001$]. Overall, choices of Option A decreased concurrent with decreases in its relative reward rates, with effectively no bias for either alternative [$b = 0.99$].

Figure 4.2b shows the proportion of choices for Option A as a function of Option A's relative reward rates during the allomorphic drug versus food procedure. Drug preferences were again under apparent control of both relative reward rates and dose, but the sensitivity to rewards rates appeared less pronounced. Results derived from the 2-D Global Exchange Matching equation confirmed the apparent trends. There was significant sensitivity to changes in relative reward rates [$S_R = 0.26$, $F(1, 64) = 18.95$, $p < .001$] and relative reward magnitude, [$S_M = 2.6$, $F(1, 64) = 251.85$, $p < .001$], and the reward magnitude for Option equated to ~2 μg/kg dose of REMI [$Ex = 2.06$]. Finally, Figure 4.3 shows the average ICIs as a function of dose for all trials as well as only the trials following REMI-Wins. Although the function showed an apparent quadratic trend, a quadratic effect of dose was not significant. ICIs significantly increased as a function of dose [$F(1, 5) = 11.4$, $p = .02$], ICIs following REMI-Wins were longer overall [$F(1, 5) = 18.34$, $p = .008$], and ICIs increased as a function of dose more rapidly following REMI-Wins than the average of all trials [$F(1, 5) = 37.43$, $p = .002$].

Molecular analysis

Similar to previous experiments, the Scaled Single-Learning model tested if altering the Base RL model's valuation equations with the exchange-rate (Ex) and sensitivity to magnitude parameters (S_M) would improve overall model fits. Indeed, inclusion of the Ex and S_M parameters resulted in overwhelming evidence in favor of relative magnitude scaling (see Table 4.1). However, unlike previous experiments, including additional α parameters in subsequent models did not result in overall improvements in model fits. According to each model's AIC value, the Scaled Single-Learning model was the best fitting model. But, upon assessing model fits it became apparent that the Scaled Single-Learning model was over fitting the 10 μg/kg and 3 μg/kg dose conditions and underfitting the 1 μg/kg condition. To assess whether the Scaled Single-Learning model was indeed underfitting the 1 μg/kg condition, each subject's average residual for the three dose conditions were plotted and compared to the next best fitting Scaled Quad-Learning model. Shown in Figure 4.4, the average residuals showed an apparent increase in the 1 μg/kg condition in four out of six subjects that was not present in the Scaled Quad-Learning; for this reason, despite the Scaled Single-Learning

model having the lower AIC, the Scaled Quad-Learning model was selected as the best-fitting model.

Figure 4.5 shows three example model fits that approximate the average AIC (± one standard deviation). Overall, the RL model tended to capture the choice dynamics of the current experiment less well than previous experiments, but still followed the general trends of the data [ΔAIC = -1190.50 relative to the chance model]. Parameters for the Scaled Quad-Learning model are shown in Figure 4.6a and are overall comparable with those from Experiment 2. However, no α values were significantly different after correcting for multiple comparisons. Finally, the calculated RPEs are illustrated in Figure 4.6b and showed the ordinally expected patterns of Option A providing greater value updating on wins than Option B as a function of dose.

Oxygen

Figures 4.7 and 4.8 show averaged oxygen traces across all animals for the most common result across measures: Dose × Option × Outcome. The average traces were concordant with ordinal predictions of RPEs. Namely, peaks tended to occur at the presentation of the win- and loss-associated stimuli while, during the reward interval, secondary or large convoluted peaks and troughs occurred for wins and losses, respectively.

For the analysis of average peaks and troughs, there were no systematic effects on troughs during the cue interval related to either Block or Dose. The proportion of trials that showed a trough during the cue interval were relatively low compared to other measures discussed below (*Mean* = 0.61, *SEM* = 0.01) and, as suggested by the average traces (Figures 4.7-4.8), cue intervals were primarily associated with peaks rather than troughs. Additionally, systematic scaling to relative reward rates was also not apparent in the current dataset. Shown in Figure 4.9, cue and reward interval peaks appeared to be trending as a function of relative reward rates when collapsing across all other variables, but changes in O_2 did not systematically scale with Block on any measure.

Oxygen peaks to the reward-associated cues were present on nearly all trials [*Mean* = 0.95, *SEM* < .01], and percent amplitude changes showed significant effects of Dose [$F(1, 511) = 3.94$, $p = .048$] and Outcome [$F(1, 511) = 6.13$, $p = .014$] and a significant Dose × Option interaction [$F(1, 511) = 4.31$, $p = .038$]. Shown in Figure

4.10a, the interaction had apparent quadratic trends in opposing directions for REMI and food, which were probed by assessing the significance of a quadratic fit. Indeed, both quadratic terms for REMI [*dose* = 0.10, $p = .333$; $dose^2 = -0.07$; $p = .048$] and food [*dose* = -0.33, $p < .001$; $dose^2 = 0.07$; $p = .023$] were significant but in opposite direction. Percent amplitude change to signals associated with food significantly decreased as a function of dose in a nonlinear fashion, while the percent amplitude changes to signals associated with REMI initially increased and then decreased as a function of dose. Additionally, the main effect of Outcome (Figure 4.9b) was due to peak amplitude changes being higher overall for cues signaling a win than those signaling a loss. Cue widths showed a significant effect of Dose [$F(1, 511) = 9.71$, $p = .002$] and significant Dose × Outcome [$F(1, 511) = 10.92$, $p = .001$] and Dose × Option × Outcome interactions [$F(1, 511) = 9.29$, $p = .002$]. Illustrated in Figure 4.10c, the interaction was driven by the average width of peaks to REMI-Win-associated cues significantly decreasing as a function of dose [*slope* = -0.07, $p < .001$], while the widths associated with REMI-Loss cues showed a non-monotonic increase and then decrease [*dose* = .06, $p = .030$; $dose^2 = -0.02$, $p = .046$].

O_2 peaks [*Mean* = 0.77, *SEM* = 0.01] and troughs [*Mean* = 0.92, *SEM* = 0.01] during the reward interval were also prevalent across trial types. The percent amplitude changes for O_2 peaks during the reward interval showed significant effects of Option [$F(1, 502) = 8.90$, $p = .003$] and Outcome [$F(1, 502) = 12.79$, $p < .001$] and significant Dose × Option [$F(1, 502) = 15.99$, $p < .001$], Option × Outcome [$F(1, 502) = 26.67$, $p < .001$], and Dose × Option × Outcome interactions [$F(1, 502) = 26.90$, $p < .001$]. Shown in Figure 4.11a, the three-way interaction was driven by Remi-Win peaks significantly increasing as a function of dose [*slope* = 0.83, $p < .001$] while no other trial outcomes showed slopes significantly different from zero. Reward-interval peak widths also corroborated the effects of amplitude. Reward-interval peak widths showed significant effects of Dose [$F(1, 502) = 18.59$, $p < .001$], Option, [$F(1, 502) = 4.79$, $p = .030$], Outcome [$F(1, 502) = 4.07$, $p = .044$], and significant Dose × Option [$F(1, 502) = 176.07$, $p < .001$], Option × Outcome [$F(1, 502) = 15.77$, $p < .001$], and Dose × Option × Outcome interactions [$F(1, 502) = 165.94$, $p < .001$]. The three-way interaction was driven by peak widths positively increasing as a function of dose [*slope* = 2.37, $p < .001$]

while no other trial outcomes showed slopes significantly different from zero (Figure 4.11b). Reward interval widths as a function of dose during REMI-Wins also showed a significant quadratic trend [*dose* = 1.66, $p < .001$, $dose^2 = 0.32$, $p = .003$], indicating widths increased nonlinearly as a function of dose.

The percent amplitude changes for O_2 troughs during the reward interval showed significant interactions of Dose × Option [$F(1, 289) = 26.82$, $p < .001$] and Dose × Outcome [$F(1, 289) = 7.68$, $p = .006$]. Both interactions showed apparent quadratic effects as a function of dose (Figure 4.12a-b). For the Dose × Option interaction, the amplitude of troughs for choices of food rewards became less negative at 3 μg/kg and then appeared to level off at 10 μg/kg [*dose* = 0.18, $p = .032$; $dose^2 = -0.07$, $p = .032$], while troughs associated with REMI choices became initially less negative at 3 μg/kg and then more negative at 10 μg/kg [*dose* = -0.53, $p < .001$]. For the Dose × Outcome interaction, troughs associated with losses became less negative at 3 μg/kg and then leveled off at 10 μg/kg [*dose* = 0.16, $p = .050$; $dose^2 = -0.07$, $p = .016$], while troughs associated with wins were less negative at 3 μg/kg and then more negative at 10 μg/kg [*dose* = -0.26, $p = .130$; $dose^2 = -0.14$, $p = .021$]. The widths of the reward-interval troughs showed significant effects of Dose [$F(1, 289) = 9.98$, $p = .002$] and Option [$F(1, 289) = 13.79$, $p < .001$] and significant Dose × Option [$F(1, 289) = 20.53$, $p < .001$], Dose × Outcome [$F(1, 289) = 31.14$, $p < .001$], and Dose × Option × Outcome interactions [$F(1, 289) = 10.61$, $p = .001$]. The three-way interaction was driven by the trough widths for REMI-Wins significantly increasing as a function of dose [*slope* = 1.11, $p < .001$], while no other conditions had slopes significantly different from zero (Figure 4.12c).

Oxygen-behavior relationships

A final set of analyses was run to assess how O_2 scaling to REMI doses related to behavioral choices scaling with REMI doses in the matching and RL models. For molar data, the ratio of peak O_2 (REMI / Food) during the reward interval was compared to the estimated reward magnitude of REMI (i.e., Remi / *Ex*) for all subjects; however, this analysis (see Figure 4.13) did not reach significance [$r = 0.46$, $p = .053$]. For molecular data, RPEs for all subjects on win trials were correlated with the percent amplitude O_2 change during the reward interval. RPEs and percent O_2 change were significantly correlated [$r = 0.24$, $p < .001$], although considerable variability was present. Finally, the

relative reward magnitude terms in the Scaled Quad-Learning RL model (i.e., λ_A / λ_B), previously expressed as the REMI dose divided by the exchange rate parameter, was replaced with the ratio of average peak O_2 (REMI / Food) obtained during the reward interval and refit to the behavioral data. Substituting peak O_2 for the behaviorally derived reward magnitude values revealed that the RL model could still successfully parameterize trial-by-trial choices. Additionally, the average AIC of the model (*AIC* = 842.40) was comparable to the other models fit with the behaviorally derived reward magnitude values and suggests feasibility in peak O2 as scaling with relative reward magnitude similar to behavioral choices.

Discussion

The purpose of the current experiment was to determine if changes in NAc O_2 activity during the allomorphic drug versus food procedure would scale with behavioral determinants of REMI's relative reward value. Similar to Experiment 2, rats' propensities to choose REMI was shown to be modulated by both the relative reward rates and magnitudes of the concurrently offered commodities. Changes in NAc O_2 activity also showed some evidence of REMI and food being scaled on a common physiological utility signal, as the appearance of reward-associated cues or reward delivery tended to result in qualitatively similar peaks of NAc O2 activity. The present experiment is believed the first to show that decisions for REMI or food produce comparable effects on NAc O_2 activity. Additionally, the data suggested that NAc O_2 activity scaled with the relative reward magnitude of the reward-associated cues, while NAc O_2 activity during the reward interval appeared to scale with the magnitude of REMI dose. However, NAc O_2 activity did not scale with relative reward rates.

With respect to behavior, the signaling condition of the current experiment was notably different than Experiment 2 in that both REMI and food choices resulted in spatially distinct win- and loss-paired cues; this was done in an attempt to achieve greater behavioral sensitivity to relative reward rates that might then increase the likelihood of observing NAc O_2 scaling. During the initial isomorphic training, the inclusion of outcome-dependent cues for food reward did indeed produce a qualitative increase in sensitivity to relative reward rates (S_R = 0.45) over Experiments 1 and 2 (avg S_R = 0.30 and 0.35, respectively). Furthermore, signaling both choice alternatives effectively

removed bias to either option. Thus, prior to exposing the rats to allomorphic drug versus food choice conditions, the rats' choices as a function of reward probability were relatively sensitive compared to previous conditions and showed very little bias for either alternative.

During the allomorphic drug vs. food procedure, the rats replicated the results of Experiment 2 by showing significant sensitivities to relative reward rates and magnitudes; these results suggest that rats making choices between REMI and food were doing so in a value-based manner. Additionally, ICIs showed a similar increase as a function of dose to those of Experiment 2 as well as another report assessing cocaine choices (Beckmann et al., 2019). However, while sensitivities to relative reward rates were significant, the qualitatively increased sensitivities seen in the isomorphic procedure returned to a comparable level as Experiment 2 during allomorphic choice. The exact reason for the decrease between conditions is unclear. One possibility could be that training rats on a descending order of REMI doses resulted in an increased REMI bias. For instance, the exchange-rate parameter (*Ex* = 2.06) for the present experiment was qualitatively lower than that of both groups in Experiment 2 (*Ex* = 3.16 and 5.86). Alternatively, it could be that the inclusion of forced-choice trials sufficiently changed the decision-making context, such that forcing rats to choose a less preferred alternative resulted in more absolute preferences. It is also interesting that there were no O_2 differences between trials of free versus forced choices, as previous work has found using voltammetry (e.g., Sugam et al., 2012). In either case, the overall results of the current experiment's molar analyses replicate the findings from Experiment 2 showing evidence of value-based decision-making for drug reward as well as independent effects of relative reward rates and magnitude predicted by matching equations (Rachlin, 1971).

Also similar to Experiment 2, RL models (Table 4.1) were successfully able to parameterize rats' allomorphic choices. Altering the valuation equations to account for relative reward magnitude scaling in the Scaled Single-Learning model again resulted in substantial improvements relative to the base RL model. Thus, for the third time, relative magnitude scaling is supported in the RL models. Additionally, the parameter estimates for the current experiment corroborated those from Experiment 2 with the exception of the perseveration parameter (*c*) being negative for the first time; this suggests that rats

had a tendency to alternate from their previous choice independently of the reward value between options, which may have been impacted by the inclusion of forced choice trials forcing rats to alternate.

For the O_2 signals, the overall results suggest that NAc activity scaled with reward magnitude for REMI infusions and possibly relative reward magnitude for the reward-associated cues. For instance, peak O_2 changes to the food-associated cues showed a nonlinear decrease as a function of dose that could be interpreted as scaling with food's relative reward magnitude. That is, food reward had the greatest preference at the 1 μg dose and then food preferences decreased as dose increased, similar to O_2 peak changes. Furthermore, given that the reward-associated cues were perfectly correlated with trial outcome, the ordinal predictions of the RPE hypothesis would suggest that the cues should produce signal changes associated with the reward magnitude of the outcome (Abler et al., 2006; Fiorillo et al., 2003; Schultz, 2010) as opposed to the probability of reward.

Alternatively, O_2 activity to REMI-associated cues showed a quadratic function not unlike that of a dose-response curve. One possible interpretation is that, similar to single-schedule measures of reward value, O_2 changes to REMI cues represent an interaction of value-based mechanisms and effects associated with REMI. For the 1 and 3 μg/kg conditions, a linear increase in NAc O_2 responses suggests possible scaling of the forthcoming REMI reward magnitude. However, at the 10 μg/kg condition, biological compensatory mechanisms may have overcome the otherwise value-based signal changes. Indeed, drug-associated cues have been shown able to produce both withdrawal symptoms and alter an individual's tolerance to drug effects, as though the body is "preparing" for drug (Siegel, 2005). Given the relatively large effect administration of the 10 μg/kg dose had on O_2 activity (discussed below), it seems likely that compensatory effects may also have been present. The interpretation that O_2 peaks to reward-associated cues scale with reward magnitude would also be consistent with work showing that individuals who are experienced with a particular drug show increased reactivity to drug-associated cues compared to inexperienced controls (e.g., David et al., 2005; Li et al., 2012). However, the interpretation is complicated in that Figure 4.10a represents O_2

responses to both REMI-Wins and REMI-Losses, making a compensatory mechanism interpretation somewhat speculative.

Win-paired cues across both options also had significantly higher peak O_2 changes. RPE hypotheses do not make specific predictions regarding win versus loss trials, but wins resulting in higher O_2 amplitude changes may reflect general excitation, which has been shown previously (e.g., Sugam et al., 2012). Finally, the widths of peak O_2 changes to reward-associated cues showed an interaction, in which REMI-Wins displayed dose-dependent decreases in widths while losses displayed an overall increase. An *a priori* reason for the dose effects on widths is not clear, but compensatory mechanisms may also be involved that resulted in a dissociation between REMI-Wins and losses.

Changes in O_2 activity during the reward interval showed large dose-dependent effects for REMI while those for food rewards were dose independent. For O_2 peaks, NAc activity for REMI-Wins during the reward interval dose-dependently increased along with increases in peak widths, but there were no changes for food delivery. These results suggest that O_2 peaks during the reward interval scaled with reward magnitude but in a nonrelative manner. That changes in NAc activity, or the broader ventral striatum area, scaled with reward magnitude manipulations corroborates (1) a large body of fMRI work (Daniel & Pollmann, 2014; Levy & Glimcher, 2012) using primarily monetary rewards, (2) electrophysiological recordings in the striatum of nonhuman primates responding for food or liquid rewards (Schultz, 2010), and (3) electrophysiological and fast-scan voltammetry recordings in the NAc of rats responding for food (e.g., Goldstein et al., 2012; Saddoris et al., 2015b). Additionally, that NAc O_2 responded in a qualitatively similar way for food and REMI rewards adds to the growing list of different reward types that the brain responds to in a similar, value-suggestive manner (Levy & Glimcher, 2012) and corroborates fast-scan voltammetry work in rats responding for cocaine (e.g., Cameron et al., 2014). However, previous work assessing brain activity relative to a commodity's reward value often do so under isomorphic conditions and, if reward magnitude is manipulated, only show the results for the commodity whose magnitude increases (e.g., Abler et al., 2006; Knutson et al., 2005). For example, the reward magnitude of food was not manipulated in the present experiment and, as such,

many studies may not have shown the activity changes in food as a function of dose; however, such changes are potentially important in determining whether brain region activity is scaling reward magnitude in a relative or independent manner. Further research is therefore necessary to corroborate the potential scaling of reward-associated cues to relative reward magnitude seen here.

Trials associated with REMI, as well as wins, also separately engendered differential O_2 trough activity. Although the three-way interaction between Option, Outcome, and Dose was not significant, O_2 trough activity was likely driven by the 10-µg REMI-Wins producing large troughs (Figure 4.7a) that is further substantiated by the significant three-way interaction for trough widths (Figure 4.12c). These troughs suggest a temporary depletion of local O_2 that then rebounded, which is consistent with rodent fMRI work that administered both REMI and morphine (Liu, Greve, Dai, Marota, & Mandeville, 2007). Furthermore, given the density of ventral striatal GABAergic medium spiny neurons that often express mu opioid receptors (Salgado & Kaplitt, 2015; Scofield et al., 2016), the 10 ug-associated troughs likely stemmed from inhibiting GABA interneurons that led to disinhibition of the NAc generally (Liu et al., 2007). While these effects of O_2 troughs likely stemmed from the direct pharmacology of REMI, they were not commensurate with RPE hypotheses. That is, reward omission should be associated with either an increased proportion or amplitude change in decreased regional activity, yet neither effect was present in the current study. However, hypotheses surrounding decreases in activity have generally been less robust and consistent (Schultz, 2010; Yacubian et al., 2006), and some would likely argue that pharmacological agents are not fair tests of RPE hypothesis. Thus, the present O_2 trough results do not present existential threats to RPE hypotheses. But, if the RPE signal ostensibly represents a common neural signal for all reward value, then all reinforcing outcomes of behavior are candidate to that pathway. It is therefore important to test the boundary conditions of our hypotheses to ensure studies do not simply repeat the same tests using small variants.

Changes in NAc activity were also assessed in relation to the observed changes in behavior at the molar and molecular level. For the molar data, the ratio of peak O_2 for REMI and food during the reward intervals were correlated with the modeled REMI and food reward magnitude ratios; this analysis showed that relative O_2 tended to covary with

the molar estimates of relative reward magnitude, but did not reach significance. For the molecular data, RL-modeled RPEs were significantly associated with peak O_2 changes as shown previously with fMRI in humans (Gläscher et al., 2010) and fast-scan voltammetry in rats (Hart et al., 2014). Additionally, replacing the reward-magnitude terms for REMI and food in the Scaled Quad-Learning RL model with relative peak O_2 ratios resulted in the RL model still able to parameterize trial-by-trial choices. Although there was considerable variability in the present measures, the correlation between peak O_2 and RPEs suggests that value updating for qualitatively different commodities (i.e., food and REMI) happens in a similar RPE-like manner within the NAc. Furthermore, while the present study found no effects of O_2 scaling with scheduled relative reward rates, the correlations between RPEs and peak O_2 changes do suggest O_2 activity was varying as a function of local rates of reward probability. Future research may therefore benefit by having increased statistical power to more effectively detect systemic effects of brain activity as a function of scheduled relative reward rates.

One possible source of the high variability in the O_2-behavior correlations may involve the use of a drug that is capable of producing direct pharmacological effects other than just the effects of reward value in the NAc. Although previous research has shown that brain measures in the striatum and prefrontal cortex for cocaine and heroin are related to behavior (e.g., Cameron et al., 2014; Guillem et al., 2018; Lenoir et al., 2007; Saddoris et al., 2016), these studies used different recording techniques and only measured one dose level relative to food without attempting to scale changes in reward magnitude with changes in brain measures. Given the ubiquity of ostensible prediction error signals found in the brain (Levy & Glimcher, 2012), future studies should scale brain and behavior changes in more dynamic procedures such as that here, which is more commonly done in the human and nonhuman primate literature.

Another possible source of variance in the present experiment may have been that the rats were freely moving (Wallis, 2012). For instance, one human fMRI study had participants either respond to a button or use eye-movement saccades to select choices and found the magnitude of the underlying neurocircuitry results were significantly different. (Wunderlich, Rangel, & O'Doherty, 2009). However, for rodents to engage in autonomous decision-making, there is little choice but to allow for freedom of movement.

Future research should therefore take the response-modality of how choices are physically made into account.

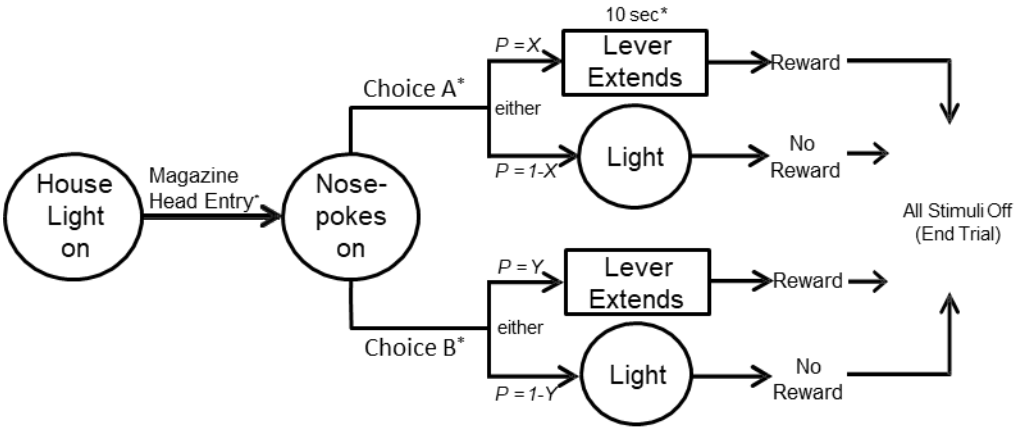
Overall, the present experiment revealed several novel findings relevant to our understanding of decision making. As in Experiment 2, rats were shown to choose as if their decisions were value-based for allomorphic choices between food rewards and REMI rewards of varying doses. Also, oxygen within the NAc was shown to respond in a way commensurate with RPE hypotheses for both food and REMI reward in freely-moving rats. Peak O_2 responses to REMI infusions were also shown to be dose-dependent and were commensurate with reward magnitude scaling. Finally, RL-modeled RPEs were shown to correlate with the peak-O_2 changes for both REMI and food rewards. The results therefore suggest that drugs may be valued on a common scale as nondrug reward, but further research is needed. Additionally, the present experiment serves as an example of how future research can increase procedural homology between preclinical studies using electrodes and human fMRI studies. Although specificity of the signal is lost as a trade-off, ever-advancing preclinical manipulations to target specific cell types such as designer drugs and receptors (Urban & Roth, 2015) or optogenetics (Fenno, Yizhar, & Deisseroth, 2011) could be coupled with future research to regain manipulative control of specific mechanisms of interest.

Table 4.1. Experiment 3 molecular models.

Model	Parameter Count	Parameters	AIC	ΔAIC
Base	*3*	α,β,c	1510.60	
Scaled Dual-Learning (Outcome)	*5*	$\alpha_{Win},\alpha_{Loss},\beta,c,S_M$	864.91	-645.69
Scaled Dual-Learning (Option)	*5*	$\alpha_{A},\alpha_{B},\beta,c,S_M$	853.96	-10.95
Scaled Quad-Learning (Option×Outcome)	*7*	$\boldsymbol{\alpha_{AWin},\alpha_{ALoss},\alpha_{BWin},\alpha_{BLoss},\beta,c,S_M}$	**809.51**	**-44.45**
Scaled Single-Learning	*4*	α,β,c,S_M	784.44	-25.07

Note: Models are shown in descending order of AIC values. ΔAIC calculations are shown as the difference from the previously best fitting model. Bolded text indicates the chosen model.

a.



b.

Probability (*P*)	X	0.66	0.22	0.11	0.11
	Y	0.11	0.11	0.22	0.66
Relative Reward Rates	A	6	2	0.50	0.17
	B	0.17	0.5	2	6

Figure 4.1. Experiment 3 general methods. (**a**) Sequence of events during the behavioral procedures on free-choice trials in Experiment 3. The reward for both options on win trials during the isomorphic procedure was one food pellet. The reward for Option A during the allomorphic procedure was an infusion of REMI (10, 3, or 1 μg/kg) and one food pellet for Option B. * indicates that all stimuli offset following the event. (**b**) Table outlining the probabilities of events and the relative reward rates for Options A and B.

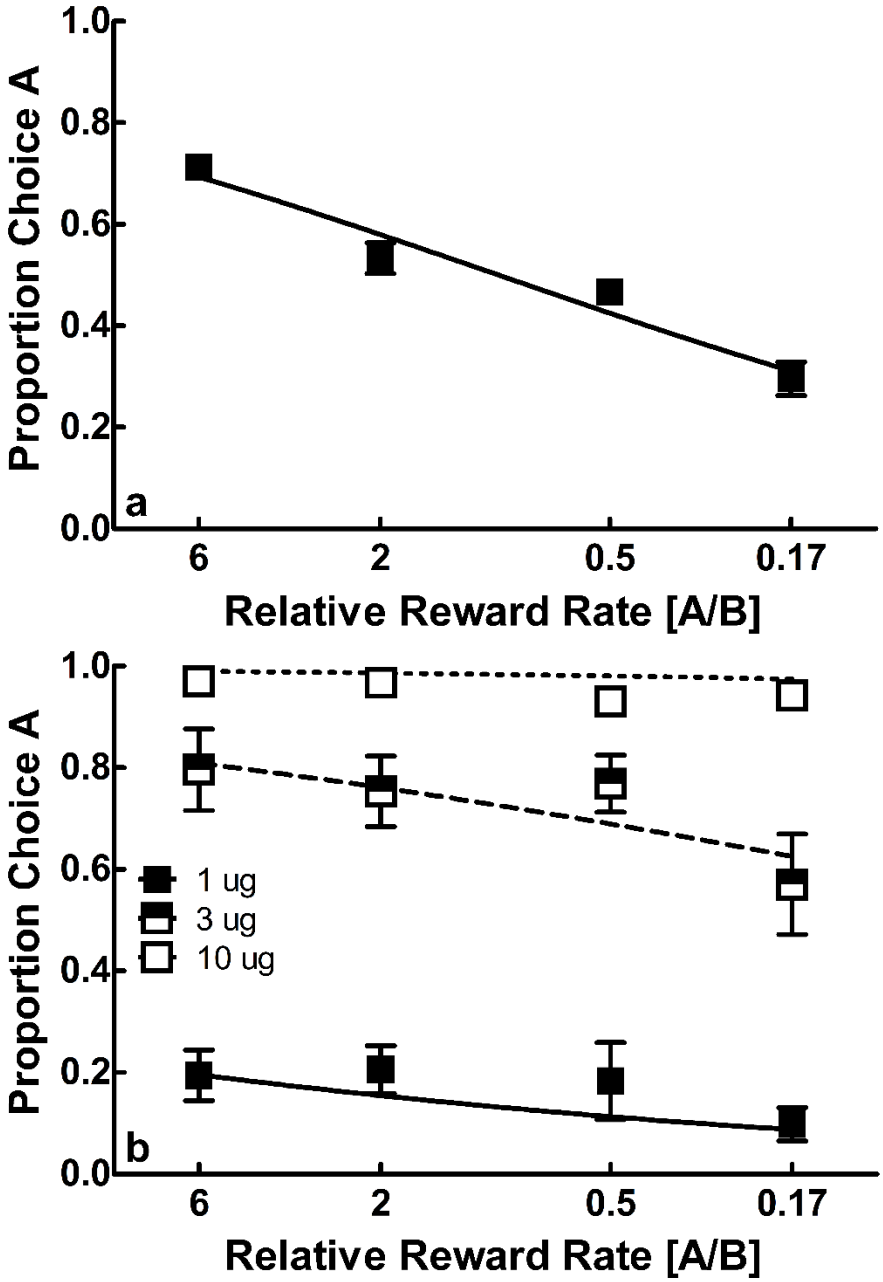


Figure 4.2. Experiment 3 molar choice data. (**a**) Mean (± SEM) proportion of choices for Option A as a function of Option A's relative reward rates during the isomorphic food versus food procedure. Lines are the NLME-determined best fit of Equation 14. (**b**) Mean (± SEM) proportion of choices for Option A as a function of Option A's relative reward rates and REMI dose during the allomorphic drug versus food procedure. Lines are the NLME-determined best fit of Equation 16. Note: x-axes are logarithmic for visualization.

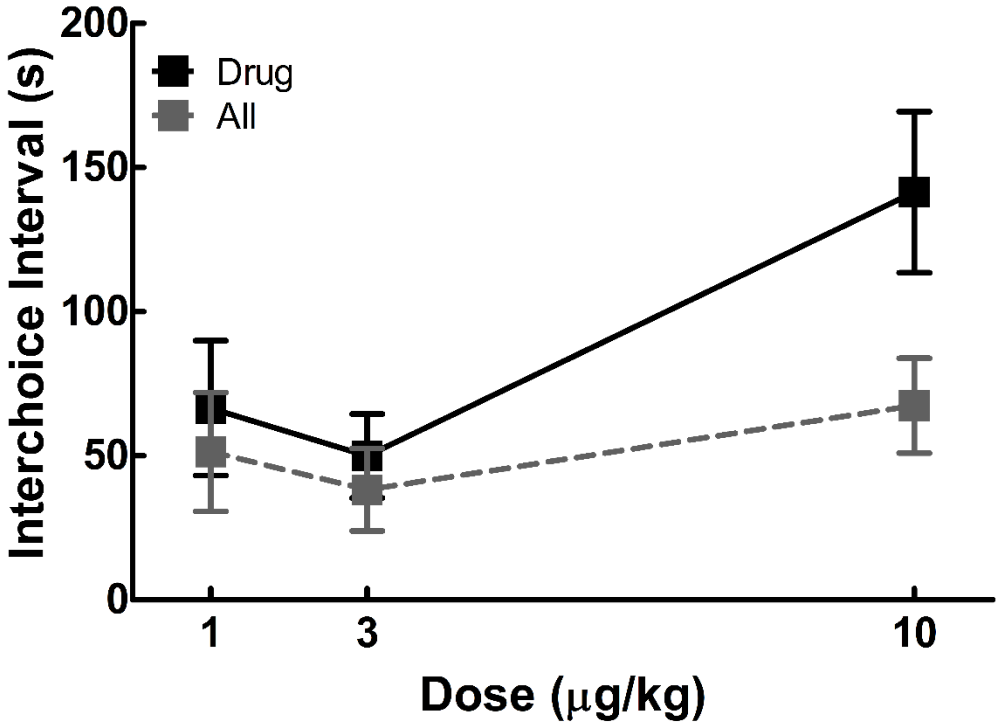


Figure 4.3. Experiment 3 interchoice intervals**.** Mean (± SEM) interchoice intervals for all trials (All) or only trials following rewarded choices for Option A (i.e., REMI-Wins; Drug) as a function of dose in Experiment 3.

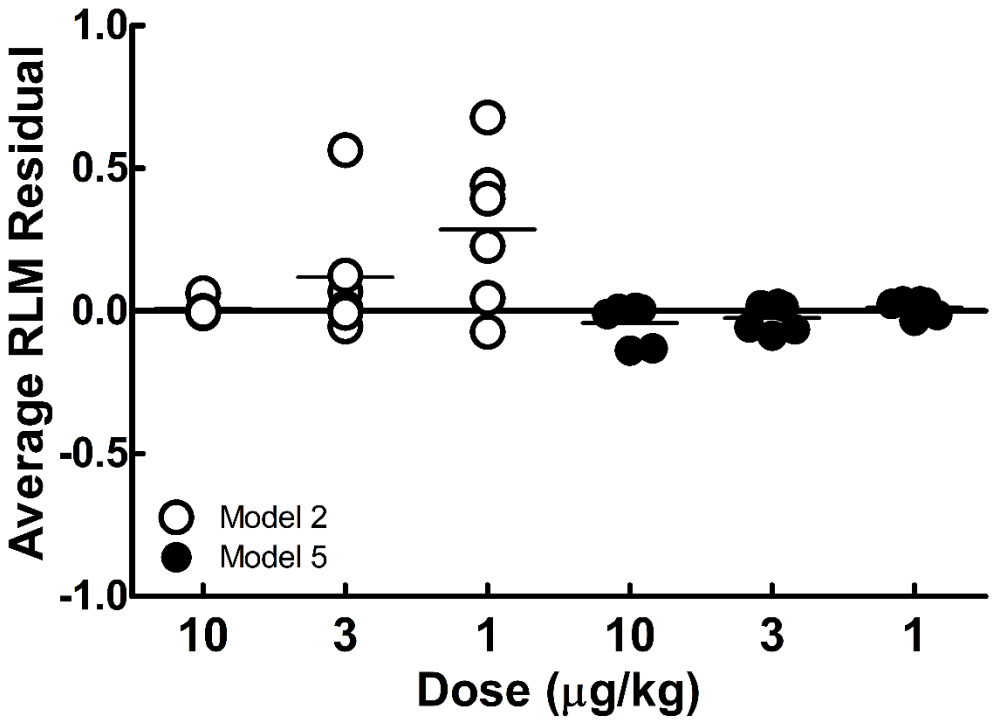


Figure 4.4. Experiment 3 RL residuals**.** Average RL residuals for the Scaled Single-Learning and Scaled Quad-Learning models as a function of dose. Data points represent the average of one individual. Horizontal lines indicate the average across individuals.

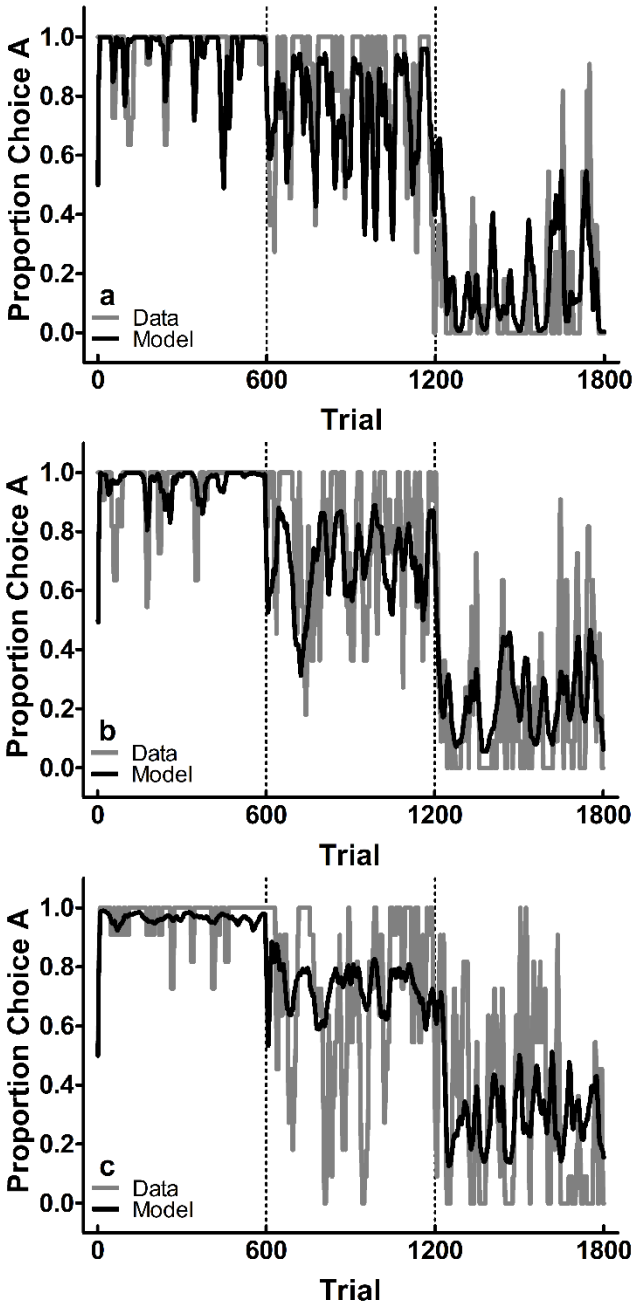


Figure 4.5. Experiment 3 RL model fits**.** Example fits from the Scaled Quad-Learning model to molecular choice data of Option A. Dashed vertical lines represent reward magnitude changes for Option A in order of 10, 3, and 1 µg/kg doses from left to right. Panels show (**a**) a rat one SD below, (**b**) a rat at, and (**c**) a rat one SD above average the AIC. All data were smoothed with an 11-trial moving average.

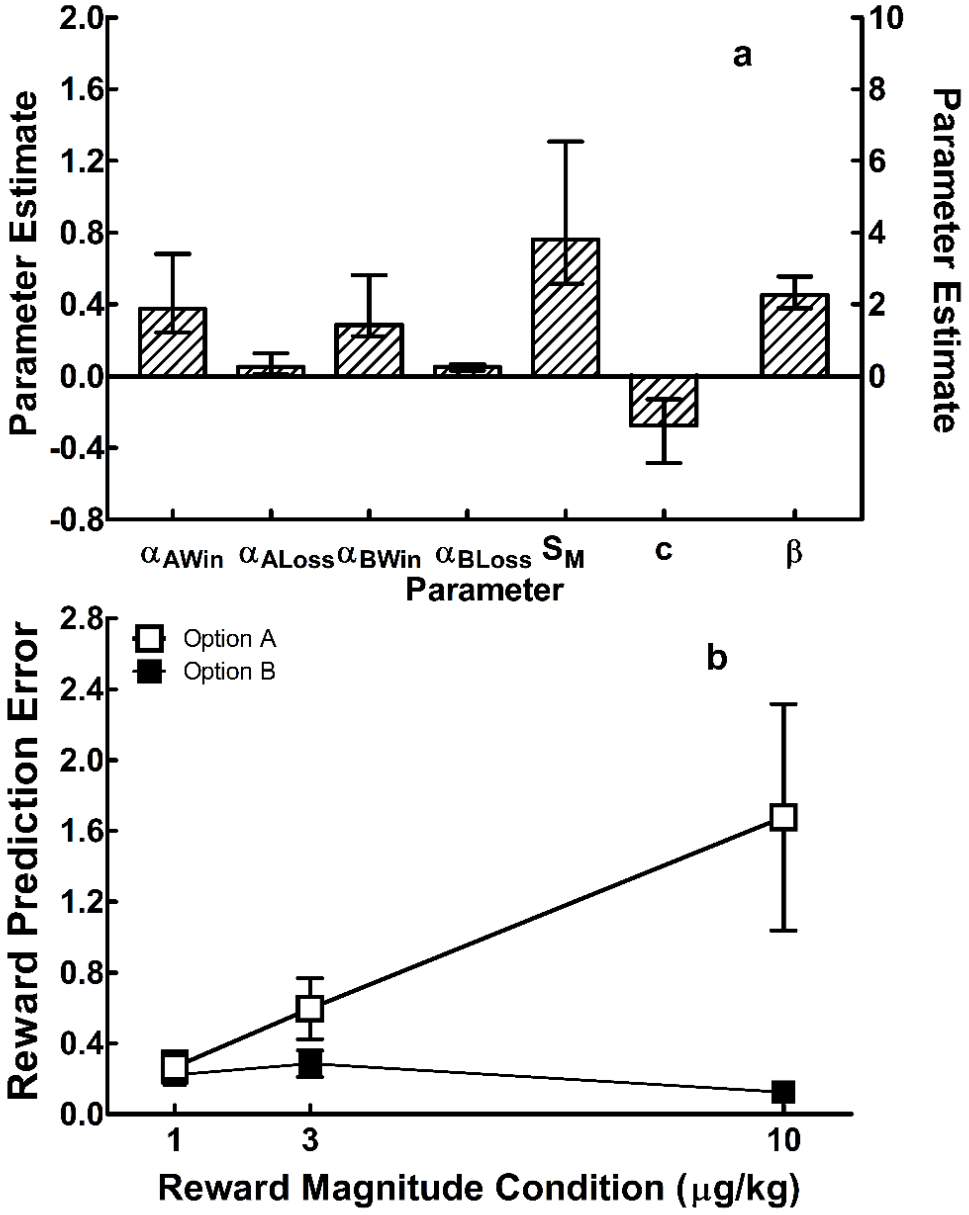


Figure 4.6. Experiment 3 molecular parameters**.** (**a**) Median parameter estimates from the Scaled Quad-Learning model fit to the molecular choice data in Experiment 3. Medians are shown due to the use of nonparametric test statistics and error bars represent the interquartile range. Note: β values are scaled according to the right y-axis. (**b**) Mean (± SEM) predicted reward prediction errors (RPEs) from the Scaled Quad-Learning model for Options A and B as a function of the reward magnitude for Option A. The displayed RPEs are from equations 18-19 when V_A and V_B equaled zero.

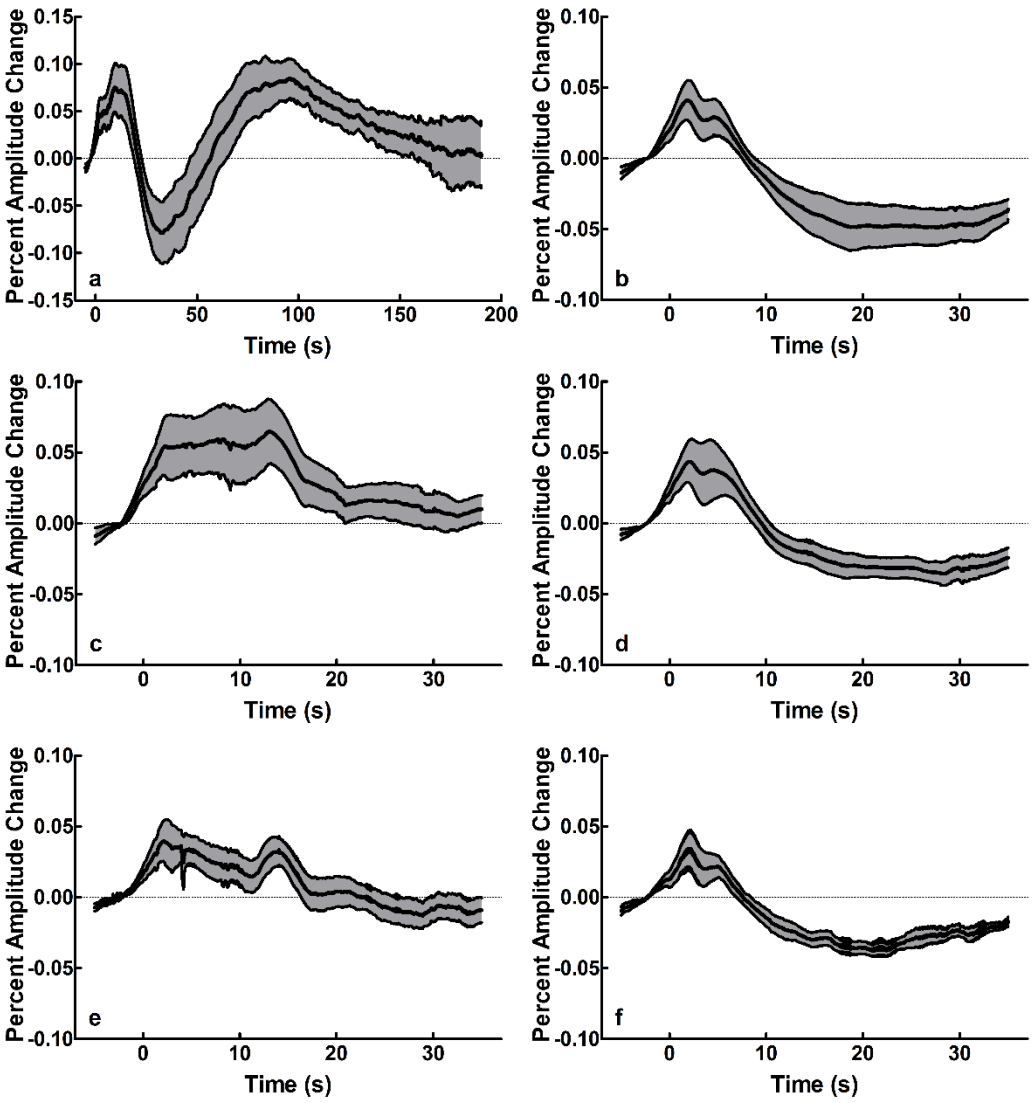


Figure 4.7. Example O_2 traces for REMI**.** Average O_2 traces for Option A (REMI) collapsing over relative reward rates and rats. (**a,c,e**) Mean (± SEM) traces for rewarded REMI choices in the 10 (**a**), 3 (**c**), and 1 (**e**) µg/kg conditions. (**b,d,f**) Mean (± SEM) traces for reward-omitted REMI choices in the 10 (**a**), 3 (**c**), and 1 (**e**) µg/kg conditions. All data were smoothed with 0.25 s moving window prior to averaging.

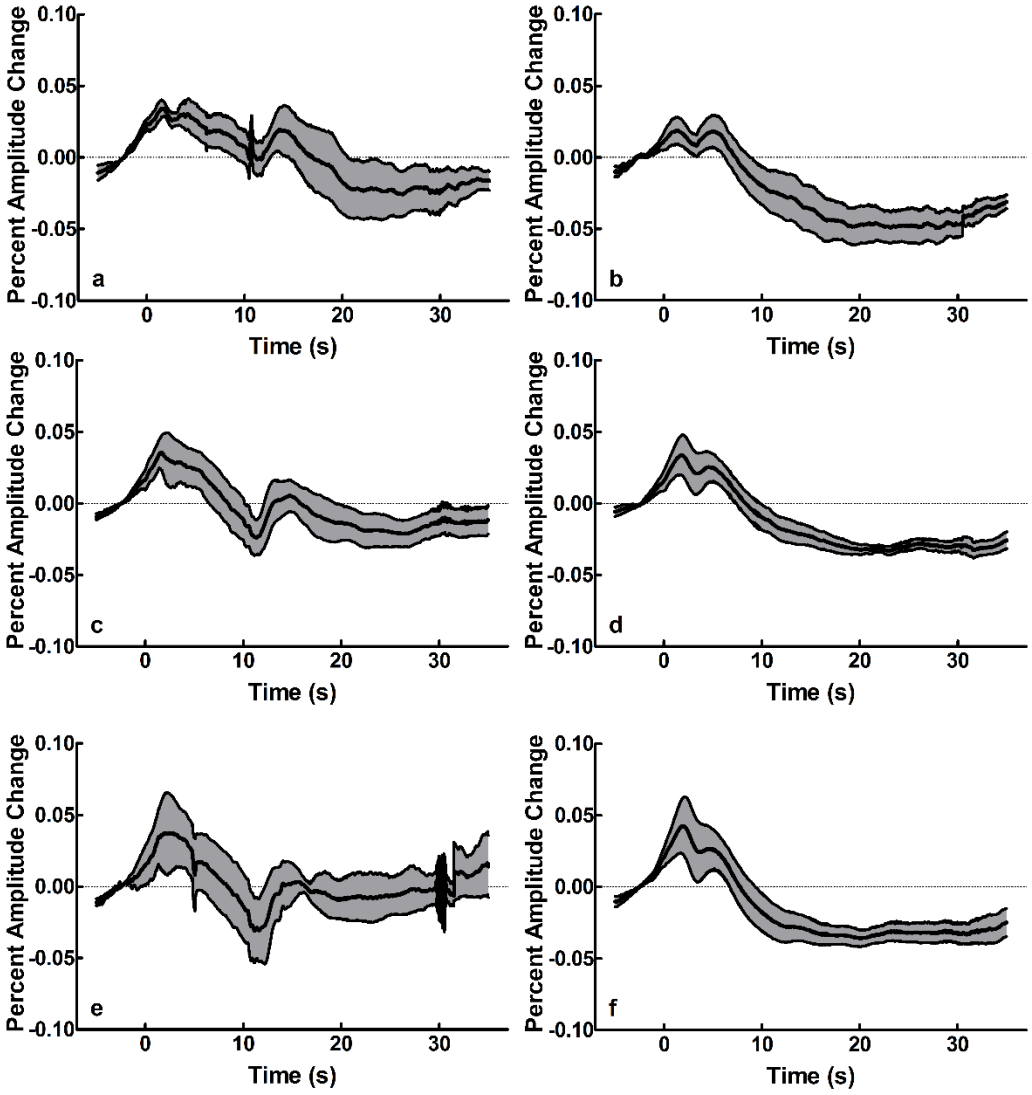


Figure 4.8. Example O_2 traces for food**.** Average O_2 traces for Option B (food) collapsing over relative reward rates and rats. (**a,c,e**) Mean (± SEM) traces for rewarded food choices in the 10 (**a**), 3 (**c**), and 1 (**e**) μg/kg conditions. (**b,d,f**) Mean (± SEM) traces for reward omitted food choices in the 10 (**a**), 3 (**c**), and 1 (**e**) μg/kg conditions. All data were smoothed with 0.25 s moving window prior to averaging.

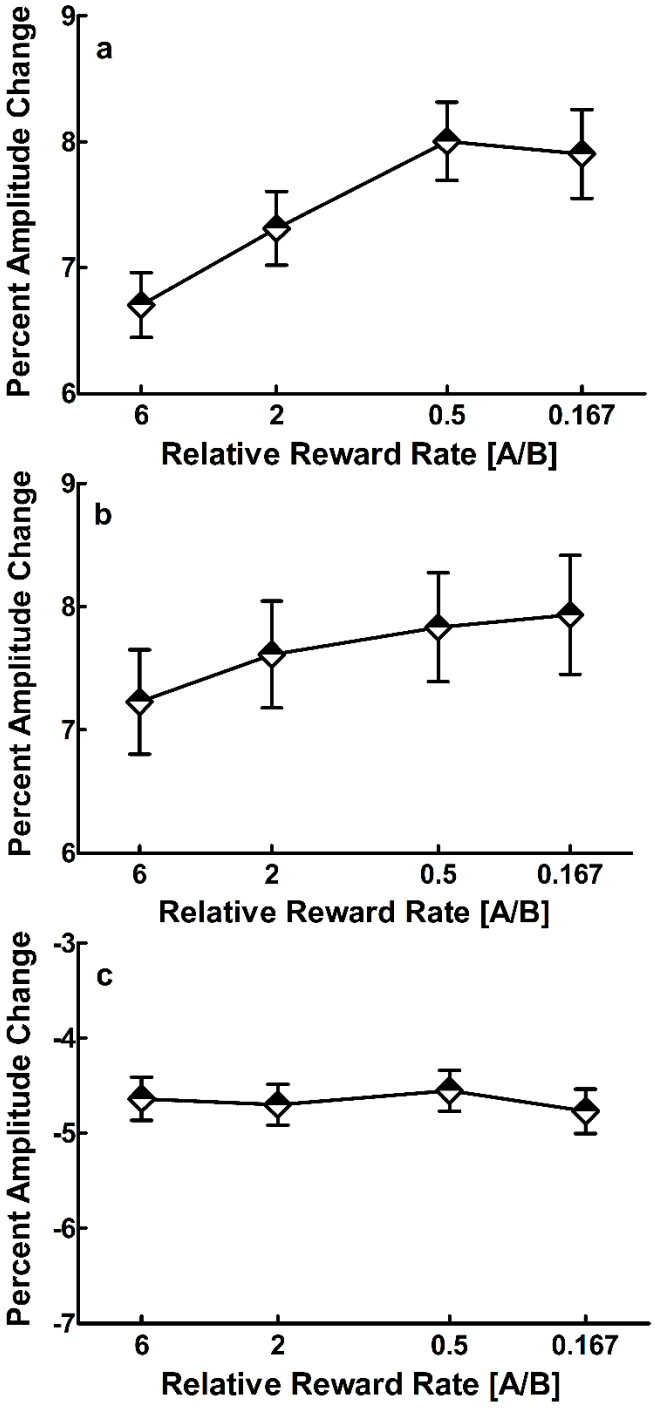


Figure 4.9. Effects of relative reward rates on NAc O_2. Mean (± SEM) O_2 percent amplitude changes for (**a**) peaks during the cue interval, (**b**) peaks during the reward interval, and (**c**) troughs during the reward interval collapsing over all other conditions. Note: x-axes are logarithmic for visualization.

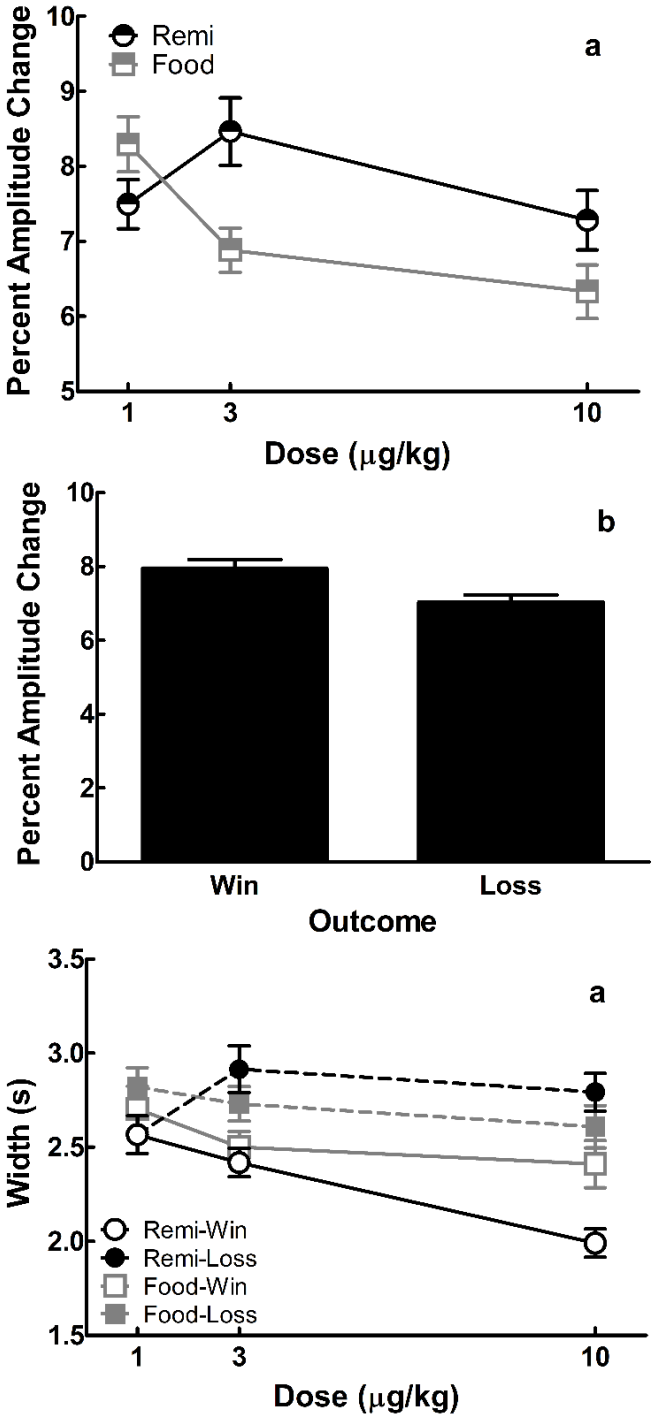


Figure 4.10. Cue interval peak O_2 effects. (**a**) Mean (± SEM) O_2 percent amplitude changes for REMI and food as a function of dose. (**b**) Mean (± SEM) O_2 percent amplitude changes as a function choice outcome. (**c**) Mean (± SEM) O_2 peak widths for REMI-Wins, REMI-Losses, Food-Wins, and Food-Losses as a function of dose.

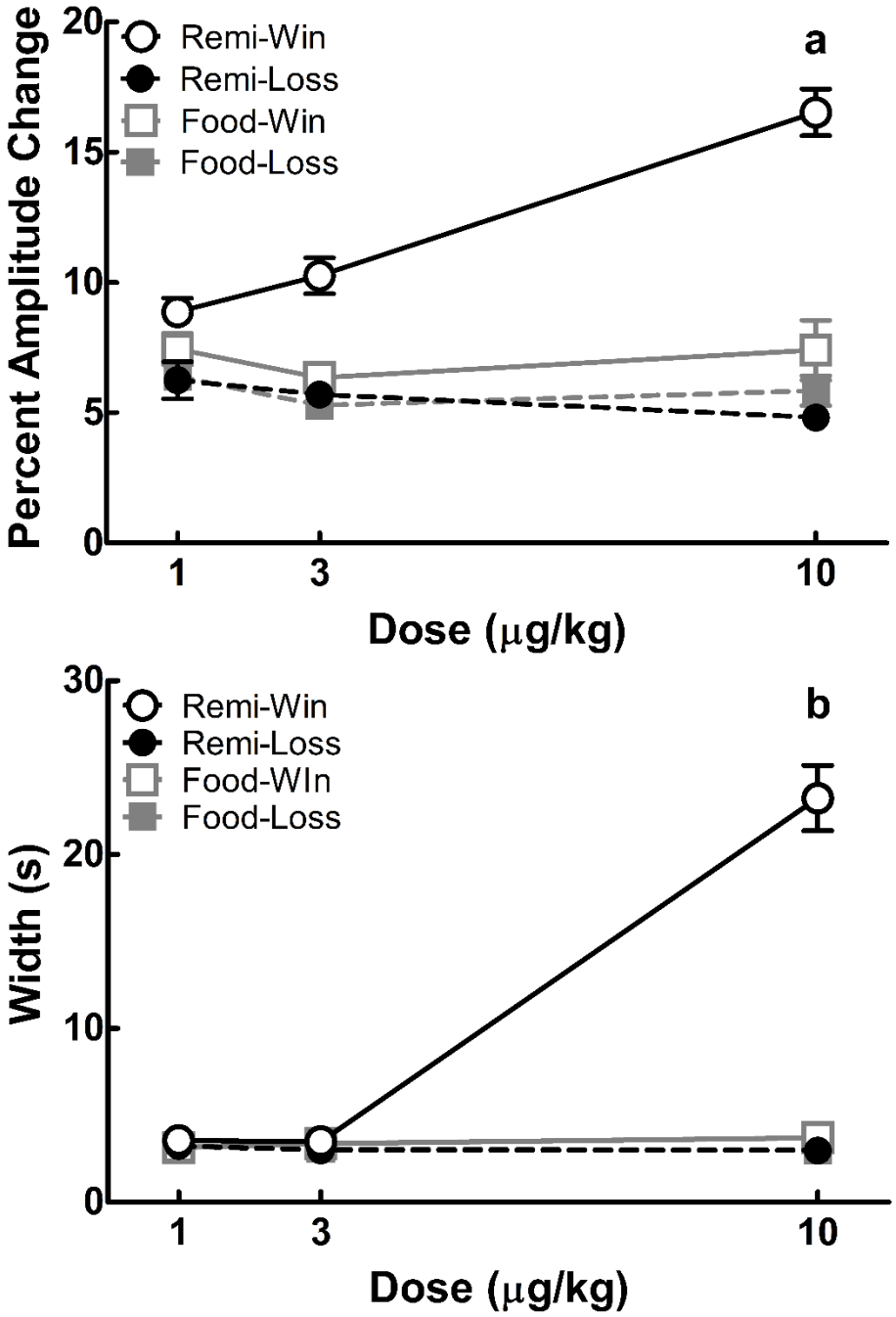


Figure 4.11. Reward interval Peak O_2 effects**.** (**a**) Mean (± SEM) O_2 percent amplitude changes for REMI-Wins, REMI-Losses, Food-Wins, and Food-Losses as a function of dose. (**b**) Mean (± SEM) O_2 peak widths for REMI-Wins, REMI-Losses, Food-Wins, and Food-Losses as a function of dose.

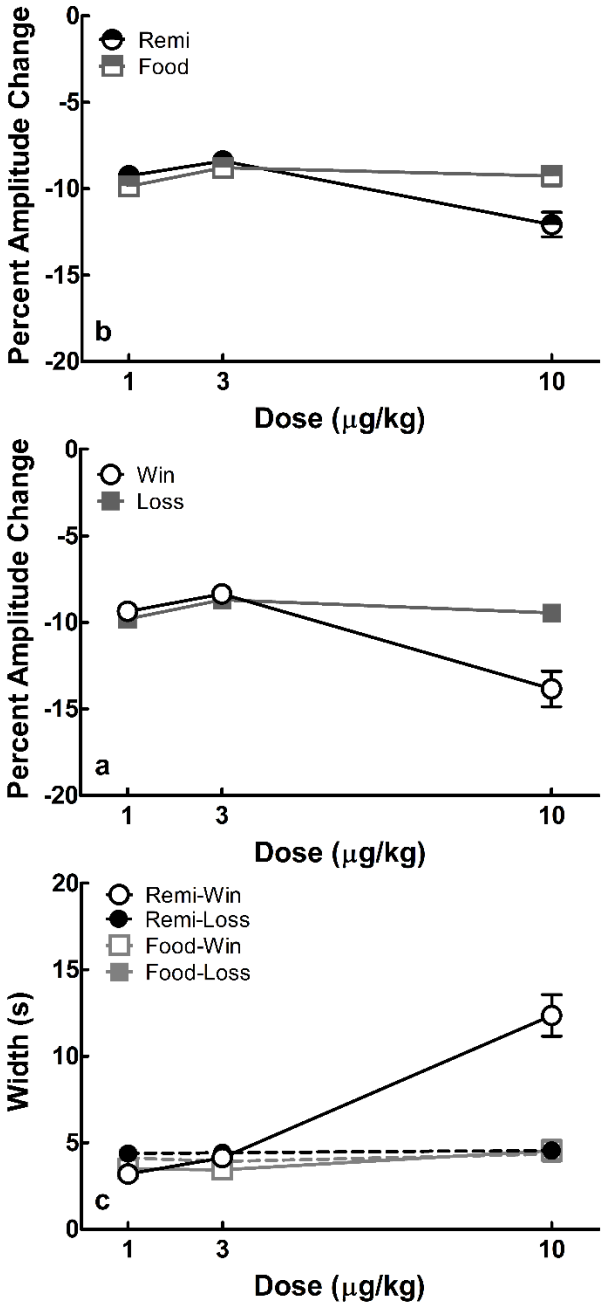


Figure 4.12. Reward interval trough O_2 effects. (**a**) Mean (± SEM) O_2 percent amplitude change for REMI and food as a function of dose. (**b**) Mean (± SEM) O_2 percent amplitude change for wins and losses as a function of dose. (**c**) Mean (± SEM) O_2 trough widths as a function of REMI-Wins, REMI-Losses, Food-Wins, and Food-Losses as a function of dose.

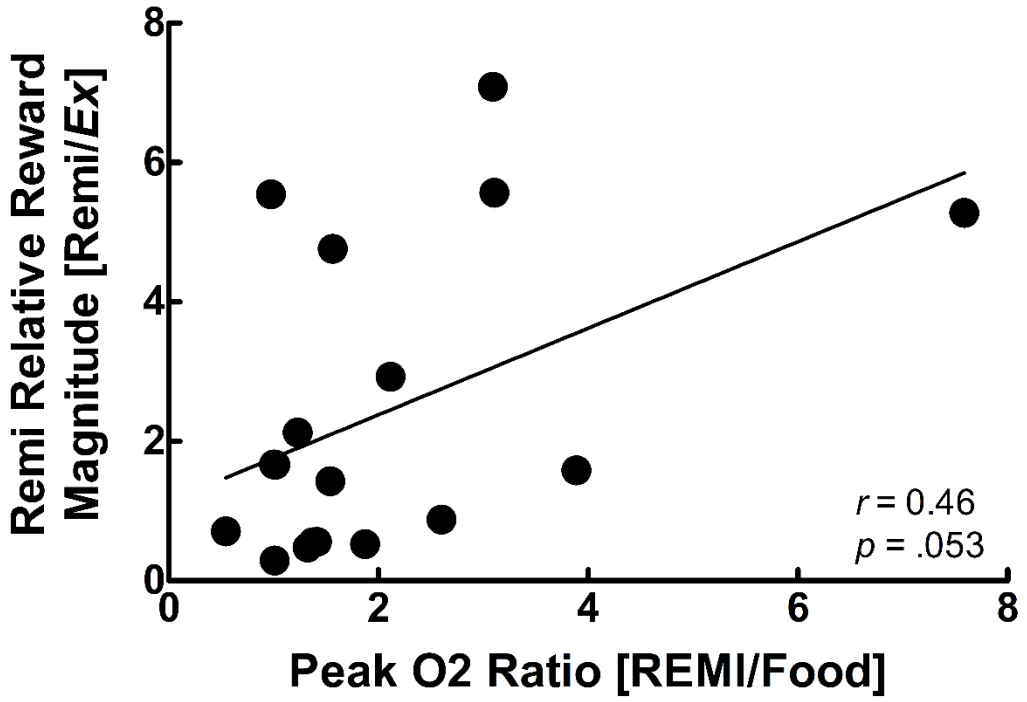


Figure 4.13. Peak O_2 correlation with modeled reward magnitudes. Correlation between REMI's behaviorally derived reward magnitude using the exchange parameter from Equation 16 and the REMI to food ratio of peak O_2 percent amplitude changes during the reward interval.

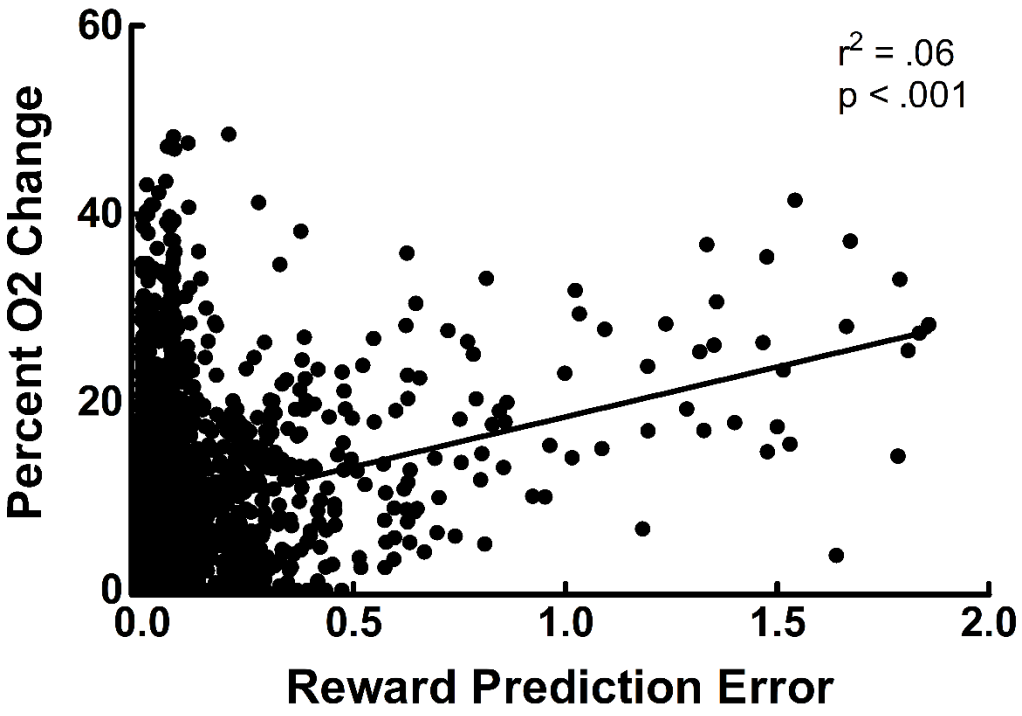


Figure 4.15. Peak O_2 correlation with modeled reward prediction errors. Correlation between the Scaled Quad-Learning reward prediction errors and the percent amplitude peak O_2 change during the reward interval for win trials.

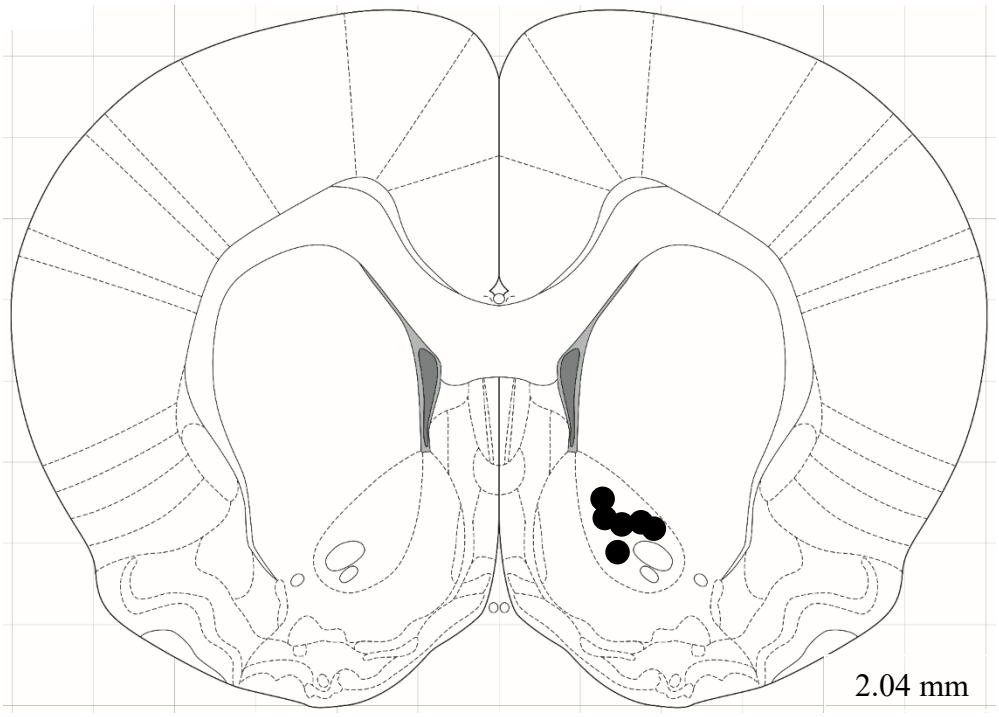


Figure 4.14. Anatomical distribution of electrode placements. Circles represent the placement of the tip of the O_2 -sensitive microelectrode arrays in the NAc (n = 6). 2.04 mm indicates the anterior position relative to bregma.

CHAPTER 5:
GENERAL DISCUSSION

The present set of experiments has benefitted our understanding of decision making in several ways. First, across all three experiments, concatenated generalized matching models (Baum, 1974; Rachlin, 1971) provided good descriptions of molar data using only one sensitivity parameter for the dimensions of relative reward rates and magnitudes. That the matching models were able capture rats' choices for both the isomorphic and allomorphic procedures substantiates the model's predictions that choices between alternatives are determined by the relative scaling of independent dimensions (Kyonka, 2008; McDowell, 2013). As such, the present experiments suggest that the use of either single schedules of reinforcement or unidimensional spaces of reward value, such as unit price, are not a complete description of choice preferences. Single schedules of reward value fail to incorporate the influence of competing commodities on reward value, while proposed unidimensional spaces do not incorporate psychophysical differences in how dimensions are scaled. Not accounting for the scaling of different reward dimensions (e.g., cost, probability, magnitude, etc.) is also a theoretically plausible reason why studies manipulating both the cost and magnitude of rewards show violations of demand theory at equivalent unit prices (e.g., Madden et al., 2000; Smith et al., 2016).

Another important finding across the current experiments was the replication of reward-associated cues biasing choice for Option A in the Signaled groups for both isomorphic food versus food and allomorphic drug versus food decision-making conditions. The use of reward-associated cues in the Signaled groups from Experiments 1 and 2 stems research done primarily with pigeons (McDevitt et al., 2016; Zentall, 2016) that has shown that cues can produce an incredible degree of insensitivity to the relative reward rates between commodities. Stated differently, the reward-associated cues created very strong biases where pigeons would repeatedly make suboptimal choices to the detriment of losing out on as much as two and a half times the food had they chosen a noncued alternative (Stagner & Zentall, 2010). Although well established in pigeons, the translation of a suboptimal choice effect to rats was a debatable effect (Chow, Smith, Wilson, Zentall, & Beckmann, 2017; Cunningham & Shahan, 2019; Smith et al., 2018;

Trujano & Orduna, 2015). The current experiments add evidence suggesting that rats' choices are indeed influenced by reward-associated cues and are also the first to show that choices for drug commodities can be influenced via the same mechanisms as choices for food rewards.

Although specifying the mechanism for why the reward-associated cues bias choice is beyond the current scope of this thesis, Experiments 1 and 2 did show that the mechanism is independent of an altered sensitivity to reward magnitude (i.e., S_M). Additionally, use of the 2-D Exchange Matching model (Equation 16) presented a novel way to quantify the effects of biased choice in terms of a commodity's relative substitutability (Green & Freed, 1993; Rachlin et al., 1976). Substitutability refers to how interchangeable two commodities are for each other. As the availability or cost of one commodity (Commodity A) changes, switching preference to another commodity (e.g., Commodity B) reflects that this latter commodity (Commodity B) acts as an economic substitute for the former (Commodity A). In the present experiments, decreases in the relative reward rates and magnitudes for Option A resulted in relative increased preferences for Option B across all experiments. As such, food was substituting for itself and serving as an economic substitute for REMI drug reward (Anderson, Velkey, & Woolverton, 2002; Beckmann et al., 2019; Koffarnus & Woods, 2008; Maguire, Gerak, & France, 2013; Negus, 2006; Schwartz et al., 2017). The use of reward-associated cues in the Signaled group also decreased the relative substitutability of Option B for Option A. That is, an unsignaled food pellet did not substitute as well for a signaled food pellet or REMI infusion relative to an unsignaled food pellet's ability to substitute for an unsignaled food pellet or REMI infusion. Overall, the results across experiments suggest that reward-associated cues change the decision-making context in a qualitative way from procedures lacking such cues.

Choice data were also analyzed at the molecular level using RL models (Sutton & Barto, 1998), which make different assumptions from matching models. RL models traditionally assume that each commodity is assigned value independent of others and is updated according to a violation between expected and obtained reward (i.e., the RPE). In all cases, formulations of RL models were successfully able to parameterize rats' decision making, including decisions for drug reward; these results suggest that rats may

choose alternatives through value-based decisions at a local, trial-by-trial level and corroborate previous uses of RL models with rodents (Groman et al., 2017; Marshall & Kirkpatrick, 2017). However, model-comparison results suggested that the assumption of independent value updating was not substantiated. Changing the reward magnitude terms into group-dependent ratios of scaled, relative reward value improved model fits to a substantial degree. Further model comparisons also revealed that the addition of parameters to allow for differential scaling of wins and losses was substantiated. That wins and losses were differentially scaled corroborates a large body of research showing the contextual-dependency of an individual's decisions (Glimcher et al., 2013; Kahneman & Tversky, 1979; Marshall & Kirkpatrick, 2017) and provides evidence that future models may need to incorporate different weighting for wins versus losses. Together, the results suggest that value updating is a relative, rather than independent, process and highlights how the testing of model assumptions can aid understanding of decision-making processes.

Finally, as shown in Experiment 3, recordings of NAc O_2 activity were generally commensurate with the possibility of a common physiological utility signal in the brain (Levy & Glimcher, 2012). NAc O_2 responded primarily as peaks to outcome-associated cues for both food and REMI, similar to previous research (e.g., Fiorillo et al., 2003; O'Doherty, Buchanan, Seymour, & Dolan, 2006). However, the present study is believed the first to use REMI as a reward commodity within a decision-making task while concurrently recording from the NAc. Experiment 3 therefore expands previous literature showing similar encoding of qualitatively different rewards (Levy & Glimcher, 2012) to also include opioid μ receptor agonist rewards. Additionally, changes in the magnitude of the peaks as a function of dose for food and REMI cues were of a manner suggestive of NAc O_2 scaling with the relative reward magnitude of food and REMI. However, definitive conclusions regarding reward-signal scaling were complicated by potential compensatory effects of cues for REMI infusions at the 10 μg/kg dose, warranting further research.

NAc O_2 activity also generally responded as peaks for both food and REMI reward delivery, but only peaks to REMI-Wins dose-dependently scaled with reward magnitude manipulations. That O_2 changes for food during the reward interval did not

vary as a function of dose may suggest that O_2 changes to reward delivery reflect scaling of reward magnitude but not in relative terms. Indeed, striatal-activity scaling with the reward magnitude of a commodity on win trials is well established (Fiallos et al., 2017; Levy & Glimcher, 2012), but methodological and/or analytical limitations often preclude determination of whether regional activity scaling of reward magnitude is relative. Future research is needed to confirm whether value in the NAc scales in relative or independent manner. Finally, peak O_2 during the reward interval significantly correlated with RL-model-derived RPEs for both food and REMI, consistent with previous research (Garrison et al., 2013). This result further corroborates the possibility that different commodity types update value according to an RPE-like rule and may guide future decisions.

Overall, the three experiments showed a relatively high degree of consistency in the behavioral results and conclusions. The present experiments employed a dynamic, probabilistic decision-making task previously used in monkeys and humans (Lau & Glimcher, 2005; Rutledge et al., 2009) and showed evidence of rats making value-based decisions by tracking the option with greater relative reward rates. Rats' preferences were also sensitive to increases in relative reward-magnitude changes, and NAc O_2 activity showed apparent sensitivity to both food- and REMI-related events that scaled with increases in REMI dose. Together, the experiments suggest isomorphic food versus food and allomorphic drug versus food choices were similarly mediated by the same value-based decision-making mechanisms. The results corroborate a large literature on economic choice theory (Davison & McCarthy, 1988; Glimcher, 2011) but also conflict with many theoretical conceptualizations of SUDs, as described next.

Applications to Substance Use Disorders

The discussion of SUDs in the preclinical literature has been largely centered around the disease model of addiction (e.g., Heather, 2017; Volkow et al., 2016). As a broad description, drug use leads to biological changes (often in mesocorticolimbic nuclei such as the NAc) that, for a variety of reasons, change the propensity of an individual to make subsequent drug choices. Taken to an extreme, several theories of addiction posit that the predicated biological changes produce insensitivities to consequences (i.e., habit; Hogarth, 2018; Vandaele & Janak, 2018) or even a compulsion to continue to choose the

drug (e.g., Robinson & Berridge, 2008; Vandaele, Cantin, Serre, Vouillac-Mendoza, & Ahmed, 2016; Wise & Koob, 2014).

The allostatic theory of addiction predicts that drug choices are initially driven by the positive reinforcement effects that drugs produce. After repeated use, the hedonic set point of an individual changes through tolerance mechanisms requiring increased drug use to both maintain positive rewarding effects of the drug and avoid withdrawal effects. Indeed, drugs can produce rewarding effects that lead to drug choices (see Experiments 2 and 3), continued drug use can lead to biological tolerance (e.g., Grecksch et al., 2006; Hooman Khademi, Farin Kamangar, Paul Brennan, & Reza Malekzadeh, 2016), and being in a withdrawal state can increase the propensity to choose drug (Negus & Rice, 2009). Thus, each of the component of the allostatic theory has evidence substantiating its claim.

However, a growing literature suggests much of the evidence for theories of addiction can be accounted for by more general value-based mechanisms. For instance, rats acutely deprived of food show increased motivation for food reward and hoarding (Morgan et al., 1943), food deprivation can produce physiological brain alterations (Blumenthal & Gold, 2010), and many humans report having experienced subjective "cravings" for different foods (Mercer & Holder, 1997). Given the evidence of the effects of food deprivation, it seems difficult to argue that the aforementioned components of the allostatic theory of addiction are unique to drug rewards. Furthermore, the changes in the brain on which many theories of addiction are predicated are correlational results (Heyman, 2013a) that do not establish a causal link between drug-induced brain changes and addiction-like behavior. Consider also that experiments using recordings of brain-region activity during decisions for food and drug (i.e., Experiment 3) increasingly show qualitatively similar functional forms of activity between the rewards (e.g., Cameron et al., 2014; Guillem et al., 2018; Levy & Glimcher, 2012). That changes in region activity are of a similar form between drug and nondrug reward is in line with the idea of a common physiological utility scale and further suggests that drug rewards may produce preferences through value-based mechanisms similar to other rewards. Indeed, the evidence from the current set of experiments supports the view that drug value was modulated in a qualitatively similar way as nondrug value.

Whereas compulsion often focuses on the positive and negative reinforcing nature of drugs of abuse, habitual behavior often includes a learning component where behavior is initially flexible and goal-directed but, with continued drug use, becomes an inflexible and automatic response to reward- or drug-associated stimuli (Everitt & Robbins, 2005; Hogarth, 2018; Robinson & Berridge, 2008; Vandaele & Janak, 2018). Prominent examples of habitual control come from studies showing that prolonged drug use will result in drug-seeking behavior despite concurrent foot shocks or presentations of shock-associated stimuli (Deroche-Gamonet, Belin, & Piazza, 2004; Vanderschuren & Everitt, 2004) or that drug rewards are not devalued by pairing with illness as food rewards are (Dickinson, Wood, & Smith, 2002). However, similar to dysregulated drug use in the escalation model (Beckmann et al., 2012), the formation of drug-associated habits also appears to be dependent on the decision-making context, as choice procedures have shown that drug choices remain goal-directed even after chronic use (Halbout, Liu, & Ostlund, 2016; Negus, 2005; Pelloux, Murray, & Everitt, 2015).

Experiments 1 and 2 also employed differential use of reward-associated cues between options in the Signaled and Unsignaled groups that could have produced inflexible habit behavior. To be in line with habit theory, rats in the Signaled group during the allomorphic choice procedure should have shown drug preferences that were insensitive to changes relative reward rates and relative doses. However, the reward-associated cues for food in Experiment 1 and drug in Experiment 2 produced the same effect of biasing choice towards the cue-paired option. Additionally, drug choices even at the 10 μg/kg condition showed significant sensitivity to changes in the relative reward rates between choice options. Suboptimal choice research alluded to previously has also shown that cues associated with food can produce robust changes in choice behavior that violate standard economic models of maximizing expected reward value (Zentall, 2016). However, in all reports of suboptimal choice to date, interpreting the cue-effects has been through a value-based manner, and removal of the cue from the decision-making context resulted in a return of individuals making choices in line with economic choice theories (Smith et al., 2017a). Thus, conditioned stimuli can certainly affect behavior, but the reward-associated cues in the present experiments biased choice for food and drug in a qualitatively similar way that suggests a more general value-based mechanism.

The idea of habitual or compulsive drug use is also challenged in human research at the laboratory and population levels. For example, substance-using populations in laboratory decision-making procedures have been shown to make decisions sensitive to the relative reward magnitude, price, and delay of both nondrug (Moeller & Stoops, 2015) and drug commodities (Griffiths, Rush, & Puhala, 1996; Jones & Comer, 2013; Moeller & Stoops, 2015). Results from the laboratory decision-making experiments are also corroborated by contingency management programs (Davis et al., 2016; Holtyn et al., 2014; Lussier, Heil, Mongeon, Badger, & Higgins, 2006) that typically offer gift or monetary vouchers contingent upon drug abstinence for a predetermined period of time. Contingency management programs are largely effective in curbing drug use, with better outcomes for programs that persist for longer or offered abstinence-contingent rewards of a larger reward magnitude. Furthermore, a number of substance users appear to remit on their own without treatment or "mature out" of dependence (Heyman, 2013a; Klingemann, Sobell, & Sobell, 2010). Indeed, perhaps the most prominent example of individuals remitting on their own for opioid use disorder comes from veterans returning from the Vietnam War (Robins, 1993). Heroin use during the Vietnam War was described as an epidemic with prevalence estimates for regular use between 20-30% of surveyed individuals. However, upon returning home, more than two thirds of the previously heroin-using veterans had no heroin use 8-10 months later. If addiction is caused by drug-induced brain changes leading to compulsive habits, it is difficult to reconcile the evidence of individuals abstaining on their own or that abstinence can be "bought" through contingency management programs (Heather, 2017; Heyman, 2013a).

The aforementioned lines of evidence conflicting with compulsive habit theories suggest a major factor when considering drug-choice propensity: the availability of alternative substitutes (Ahmed, 2005; Beckmann et al., 2019). When the only option available to an individual is to take drug as in single schedules, behavioral symptoms congruent with compulsive habits emerge. When another alternative commodity that can compete with drug reward is available, compulsive behavior is often absent and drug preferences depend upon the relative position of each reward along many dimensions of reward value (i.e., concatenated matching).

The importance of choice preferences being relative, while seemingly simple, cannot be understated. As just described, many papers have suggested that drug reward results in compulsive and habitual behavior due to drugs' inherently superior reward value. The opposite end of the spectrum also exists. Some research groups have suggested that drug rewards are in fact inherently inferior (Ahmed, Lenoir, & Guillem, 2013; Cantin et al., 2010; Lenoir et al., 2007; Venniro et al., 2018) due to being unable to show preference reversals as a function of increases in the drug dose offered during choice procedures. However, neither drug nor nondrug rewards are inherently better than the other. Preferences depend on the decision-making context, and the dose of the drug (i.e., reward magnitude) is only one of many operable dimensions that can modulate reward value. For instance, consider the allomorphic choice results of Experiment 2 (Figure 3.1b-c). One would have a particularly hard time trying to describe whether REMI or food was the more rewarding commodity. If the REMI dose offered is 10 μg/kg, REMI is largely preferred, but if the REMI dose is 1 μg/kg, food is largely preferred. In a similar way, a 3-μg/kg infusion of REMI is weakly preferred to food when it is paired with a reward-associated stimulus with favorable reward rates but not when the reward rates favor food and never when the infusion is unsignaled. Although failure to produce dose-dependent drug preferences (as described above) was surprising, concluding a commodity-superiority hierarchy off the basis of so few dimensional manipulations may be inappropriate. Had Experiment 2 only tested the Unsignaled group using 1 and 3 μg/kg infusions, a conclusion of drugs being inferior rewards might also have been satisfied. Instead, future research should test many points along many reward dimensions (e.g., delay, effort, probability, magnitude).

Changing the Conversation About SUDs

The conceptualization of SUDs in the present thesis has largely been centered on conflicting hypotheses that SUDs are driven by a manifestation of biologically-mediated compulsion or by a disorder of choice. Drug choice in the present experiments showed no evidence of compulsion, and the present results corroborate a larger body of work advocating that SUDs may be well described as a disorder of choice. However, biological mechanisms mediating SUDs are still critical determinants of such behaviors. For example, physical dependence and withdrawal are important risk factors that can change

an individual's propensity to make drug choices (Negus & Rice, 2009; Williams et al., 2001). Additionally, pharmacotherapies, such as agonist replacement therapies, have shown efficacy in reducing drug choices (Kampman & Jarvis, 2015; Rush & Stoops, 2012). Thus, SUDs should continue to be discussed as an interaction between decision-making mechanisms and biological mechanisms.

Although biological variables are important and can influence drug choices, compulsive drug choice does not seem a useful description that will facilitate future research or treatment approaches (Heather, 2017; Heyman, 2013a; Hogarth, 2018). For instance, the meaning of "compulsive choice" is somewhat vacuous, as it is not used consistently within the field (Heather, 2017) and often appears to act as a synonym for "strong preference" (Heyman, 2003). Some may argue that the debate of word choice is merely semantics, but the idea of a strong preference has a place within choice and decision-making theories whereas compulsions currently do not. Indeed, reframing words like compulsion or craving into a state-dependent bias for a commodity has found success in studies attempting to quantify the effects of "craving" on choice preferences (Konova, Louie, & Glimcher, 2018). Using different terms does not change the complexity of SUDs, but it may facilitate future research by speaking on more common grounds that could be supported by mathematical choice models.

A final point to consider is what the characterization of SUDs as a disorder of choice might mean at a policy level. For instance, some may disagree with the idea of conceptualizing SUDs as a disorder of choice, not for scientific reasons, but because calling drug use a choice may implicate a social stigma such that individuals would not be deserving of help or treatment (Heather, 2017). Indeed, social stigmatization represents a pertinent issue for drug users (Lloyd, 2013), but it is also one that is already present in our society. Furthermore, although a disease model characterization of SUDs may exonerate the individual of blame to some degree, it may also remove the locus of control from the individual and generate a feeling of powerlessness to change (Heather, 2017).

A choice model of SUDs also does not have to mean that drug choices are failures of self-control or volitional will power. Some posited choice theories are able to describe how an individual may behave *as if* they are compulsively choosing drug but are actually

maximizing utility (Heyman, 2013b; Rachlin, 1997). As previously described, these choice theories dissociate between maximizing local utility (i.e., at the present moment) versus global utility (i.e., longer term utility). If a drug of abuse is chosen, then one assumes that the drug produced greater local utility relative to other available substitutes at the time of choice. As such, an individual maximizing local utility will myopically continue to choose drug reward until a better alternative becomes available. However, with continued use, the repeated drug choices decrease its own subjective utility (due to tolerance) but may also decrease the utility of other potential substitutes. For instance, drug choices can preclude the availability of future substitutes through social isolation, job loss, incarceration, and declining health. Additionally, the benefits of choosing drug reward are immediate, while both drug-associated costs and many alternative rewards have delayed and uncertain outcomes that results in a decreased impact on behavior. Thus, although staying healthy, spending time with friends and family, or maintaining a job might produce overall greater utility, at the time of choice, drug reward produces greater local utility. As such, it is possible for an individual to choose *as if* they are compulsively seeking drug but may instead be maximizing utility without invoking ideas about compulsion or willpower. Alternatively, a compulsive individual should not show value-based decision making as in the present experiments and may suggest a value-based conceptualization of SUDs offers a viable alternative.

Another main point from the above example is that the availability of economic substitutes for drug reward is one possible mechanism to decrease drug choice and is consistent with the results of the current experiments. Availability of substitutes may also be a moderating factor of drug use at a systemic level within society. For example, preclinical models that provide rats with an enriching environment full of toys, conspecifics, and access to running wheels can reduce drug choices (Alexander, Coambs, & Hadaway, 1978; Bardo, Klebaur, Valone, & Deaton, 2001; Hofford, Chow, Beckmann, & Bardo, 2017), while rats in impoverished conditions lacking the available substitutes choose drug more often. Although not a perfect metaphor, the effects of enrichment certainly seem to parallel the findings that individuals of lower socioeconomic status are more prone to drug use (Galea & Vlahov, 2002; Williams & Latkin, 2007) and may reflect a lack of available alternatives to drug rewards. Clearly, decisions on how to

conceptualize SUDs are complicated, but avoiding a choice model for fear of a societally placed stigma may also preclude discussion leading to potential treatments.

Conclusions

The purpose of the present experiments was to assess whether predictions of economic choice theory were upheld in an allomorphic drug versus food choice procedure. Namely, the predictions were that choices for drug reward would be sensitive to changes in relative reward rates and magnitudes. In Experiment 1, the isomorphic food versus food procedure was shown to provide a valid preclinical model of probabilistic decision making in rats, in which choices were sensitive to the relative reward rates and magnitudes of concurrently available options. For allomorphic drug versus food decision making in Experiment 2, the results did not change. Rats' choices between the opioid μ receptor agonist remifentanil and food still showed sensitivity to relative reward rates and magnitudes, suggesting that rats chose as if through value-based decision-making processes. Finally, Experiment 3 was generally commensurate the possibility of a common physiological reward signal between REMI and food within the NAc. Namely, both reward types showed similar peaks to reward-related events and Nac O_2 dose-dependently scaled with REMI infusions. The results have potentially broad implications for current conceptualizations of SUDs, including a potential need to change the way SUDs are discussed and modeled in preclinical experiments. Overall, the present set of experiments provides additional evidence that a choice model of SUDs may be a viable alternative to the disease model and could provide insight into novel treatments based on economic principles of reward value.

REFERENCES

Abler, B., Walter, H., Erk, S., Kammerer, H., & Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage, 31*(2), 790-795.

Ahmed, S. H. (2005). Imbalance between drug and non-drug reward availability: a major risk factor for addiction. *European journal of pharmacology, 526*(1-3), 9-20.

Ahmed, S. H. (2010). Validation crisis in animal models of drug addiction: Beyond non-disordered drug use toward drug addiction. *Neuroscience & Biobehavioral Reviews, 35*(2), 172-184. doi:http://dx.doi.org/10.1016/j.neubiorev.2010.04.005

Ahmed, S. H., & Koob, G. (1998). Transition from moderate to excessive drug intake: change in hedonic set point. *Science, 282*(5387), 298-300.

Ahmed, S. H., Lenoir, M., & Guillem, K. (2013). Neurobiology of addiction versus drug use driven by lack of choice. *Current Opinion in Neurobiology, 23*(4), 581-587.

Ahmed, S. H., Walker, J. R., & Koob, G. F. (2000). Persistent increase in the motivation to take heroin in rats with a history of drug escalation. *Neuropsychopharmacology, 22*(4), 413.

Alessi, S. M., & Petry, N. M. (2003). Pathological gambling severity is associated with impulsivity in a delay discounting procedure. *Behavioural Processes, 64*(3), 345-354.

Alexander, B. K., Coambs, R. B., & Hadaway, P. F. (1978). The effect of housing and gender on morphine self-administration in rats. *Psychopharmacology, 58*(2), 175-179.

Anderson, K. G., Velkey, A. J., & Woolverton, W. L. (2002). The generalized matching law as a predictor of choice between cocaine and food in rhesus monkeys. *Psychopharmacology, 163*(3-4), 319-326.

Antoniou, K., Kafetzopoulos, E., Papadopoulou-Daifoti, Z., Hyphantis, T., & Marselos, M. (1998). D-amphetamine, cocaine and caffeine: a comparative study of acute effects on locomotor activity and behavioural patterns in rats. *Neuroscience & Biobehavioral Reviews, 23*(2), 189-196.

Apicella, P., Ljungberg, T., Scarnati, E., & Schultz, W. (1991). Responses to reward in monkey dorsal and ventral striatum. *Experimental brain research, 85*(3), 491-500.

Arvanitogiannis, A., & Shizgal, P. (2008). The reinforcement mountain: allocation of behavior as a function of the rate and intensity of rewarding brain stimulation. *Behavioral Neuroscience, 122*(5), 1126.

Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington, DC: Author.

Attwell, D., & Iadecola, C. (2002). The neural basis of functional brain imaging signals. *Trends in Neurosciences, 25*(12), 621-625.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism, 21*(10), 1133-1145.

Ballard, K., & Knutson, B. (2009). Dissociable neural representations of future reward magnitude and delay during temporal discounting. *NeuroImage, 45*(1), 143-150. doi:10.1016/j.neuroimage.2008.11.004

Banks, M. L., & Negus, S. S. (2012). Preclinical Determinants of Drug Choice under Concurrent Schedules of Drug Self-Administration. *Advances in pharmacological sciences, 2012*, 281768-281768. doi:10.1155/2012/281768

Bardo, M., & Bevins, R. A. (2000). Conditioned place preference: what does it add to our preclinical understanding of drug reward? *Psychopharmacology, 153*(1), 31-43.

Bardo, M., Klebaur, J., Valone, J., & Deaton, C. (2001). Environmental enrichment decreases intravenous self-administration of amphetamine in female and male rats. *Psychopharmacology, 155*(3), 278-284.

Barrus, M. M., Cherkasova, M., & Winstanley, C. A. (2015). Skewed by Cues? The Motivational Role of Audiovisual Stimuli in Modelling Substance Use and Gambling Disorders. In *Behavioral Neuroscience of Motivation* (pp. 507-529): Springer.

Baum, W. M. (1974). On two types of deviation from the matching law: bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22*(1), 231-242.

Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior, 12*(6), 861-874.

Baylon, G. J., Kaplan, H. L., Somer, G., Busto, U. E., & Sellers, E. M. (2000). Comparative abuse liability of intravenously administered remifentanil and fentanyl. *Journal of clinical psychopharmacology, 20*(6), 597-606.

Beauchaine, T. P., Ben-David, I., & Sela, A. (2017). Attention-deficit/hyperactivity disorder, delay discounting, and risky financial behaviors: A preliminary analysis of self-report data. *PLoS ONE, 12*(5), e0176933.

Beckmann, J. S., Chow, J. J., & Hutsell, B. A. (2019). Cocaine-associated decision-making: Toward isolating preference. *Neuropharmacology*.

Beckmann, J. S., Gipson, C. D., Marusich, J. A., & Bardo, M. T. (2012). Escalation of cocaine intake with extended access in rats: dysregulated addiction or regulated acquisition? *Psychopharmacology, 222*(2), 257-267.

Belin-Rauscent, A., Fouyssac, M., Bonci, A., & Belin, D. (2016). How preclinical models evolved to resemble the diagnostic criteria of drug addiction. *Biological Psychiatry, 79*(1), 39-46.

Bentzley, B. S., Fender, K. M., & Aston-Jones, G. (2013). The behavioral economics of drug self-administration: a review and new analytical approach for within-session procedures. *Psychopharmacology, 226*(1), 113-125.

Bentzley, B. S., Jhou, T. C., & Aston-Jones, G. (2014). Economic demand predicts addiction-like behavior and therapeutic efficacy of oxytocin in the rat. *Proceedings of the National Academy of Sciences, 111*(32), 11822-11827.

Bernoulli, D. (1738/1954). Exposition of a new theory on the measurement of risk [translation by L. Sommer of D. Bernoulli, 1738, Specimen theoriae novae de mensura sortis, *Papers of the Imperial Academy of Science of Saint Petersburg 5*, 175-192]. *Econometrica, 22*(1), 23-36.

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology, 9*(1), 65-73. doi:10.1016/j.coph.2008.12.014

Bickel, W. K., Johnson, M. W., Koffarnus, M. N., MacKillop, J., & Murphy, J. G. (2014). The behavioral economics of substance use disorders: reinforcement pathologies and their repair. *Annual review of clinical psychology, 10*, 641-677.

Bickel, W. K., Marsch, L. A., & Carroll, M. E. (2000). Deconstructing relative reinforcing efficacy and situating the measures of pharmacological reinforcement with behavioral economics: a theoretical proposal. *Psychopharmacology, 153*(1), 44-56.

Birnbaum, M. H. (1992). Issues in utility measurement. *Organizational Behavior and Human Decision Processes, 52*(3), 319-330.

Blanchard, T. C., Hayden, B. Y., & Bromberg-Martin, E. S. (2015). Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron, 85*(3), 602-614.

Blumenthal, D. M., & Gold, M. S. (2010). Neurobiology of food addiction. *Current Opinion in Clinical Nutrition & Metabolic Care, 13*(4), 359-365.

Bonem, M., & Crossman, E. K. (1988). Elucidating the effects of reinforcement magnitude. *Psychological Bulletin, 104*(3), 348-362. doi:10.1037/0033-2909.104.3.348

Branch, M. N. (1984). Rate dependency, behavioral mechanisms, and behavioral pharmacology. *Journal of the Experimental Analysis of Behavior, 42*(3), 511-522.

Browne, R. G., & Segal, D. S. (1980). Behavioral activating effects of opiates and opioid peptides. *Biological Psychiatry*.

Bulley, A., & Gullo, M. J. (2017). The influence of episodic foresight on delay discounting and demand for alcohol. *Addictive Behaviors, 66*, 1-6.

Bush, R. R., & Mosteller, F. (1953). A stochastic model with applications to learning. *The Annals of Mathematical Statistics*, 559-585.

Buxton, R. B. (2012). Dynamic models of BOLD contrast. *NeuroImage, 62*(2), 953-961.

Caine, S. B., & Koob, G. F. (1994). Effects of dopamine D-1 and D-2 antagonists on cocaine self-administration under different schedules of reinforcement in the rat. *Journal of Pharmacology and Experimental Therapeutics, 270*(1), 209-218.

Cameron, C. M., Wightman, R. M., & Carelli, R. M. (2014). Dynamics of rapid dopamine release in the nucleus accumbens during goal-directed behaviors for cocaine versus natural rewards. *Neuropharmacology, 86*, 319-328.

Cantin, L., Lenoir, M., Augier, E., Vanhille, N., Dubreucq, S., Serre, F., . . . Ahmed, S. H. (2010). Cocaine is low on the value ladder of rats: possible evidence for resilience to addiction. *PLoS ONE, 5*(7), e11592.

Caprioli, D., Zeric, T., Thorndike, E. B., & Venniro, M. (2015). Persistent palatable food preference in rats with a history of limited and extended access to methamphetamine self-administration. *Addiction Biology, 20*(5), 913-926.

Carlson, R. G., Nahhas, R. W., Martins, S. S., & Daniulaityte, R. (2016). Predictors of transition to heroin use among initially non-opioid dependent illicit pharmaceutical opioid users: A natural history study. *Drug and Alcohol Dependence, 160*, 127-134.

Carroll, M. E., & Rodefer, J. S. (1993). Income alters choice between drug and an alternative nondrug reinforcer in monkeys. *Experimental and Clinical Psychopharmacology, 1*(1-4), 110.

Case, J. P., & Zentall, T. R. (2018). Suboptimal choice in pigeons: Does the predictive value of the conditioned reinforcer alone determine choice? *Behavioural Processes, 157*, 320-326.

Catania, A. C. (1963a). Concurrent performances: A baseline for the study of reinforcement magnitude. *Journal of the Experimental Analysis of Behavior, 6*(2), 299-300.

Catania, A. C. (1963b). Concurrent performances: Reinforcement interaction and response independence. *Journal of the Experimental Analysis of Behavior, 6*(2), 253-263.

CDC. (2017). *Understanding the Epidemic*. Retrieved from National Center for Injury Prevention and Control, Division of Unintentional Injury Prevention.: https://www.cdc.gov/drugoverdose/epidemic/index.html

Childs, E., & de Wit, H. (2009). Amphetamine-induced place preference in humans. *Biological Psychiatry, 65*(10), 900-904.

Chow, J. J., Smith, A. P., Wilson, A. G., Zentall, T. R., & Beckmann, J. S. (2017). Suboptimal choice in rats: Incentive salience attribution promotes maladaptive decision-making. *Behavioural Brain Research, 320*, 244-254. doi:http://dx.doi.org/10.1016/j.bbr.2016.12.013

Christensen, C. J., Silberberg, A., Hursh, S. R., Huntsberry, M. E., & Riley, A. L. (2008). Essential value of cocaine and food in rats: tests of the exponential model of demand. *Psychopharmacology, 198*(2), 221-229.

Crespo, J. A., Sturm, K., Saria, A., & Zernig, G. (2005). Simultaneous intra-accumbens remifentanil and dopamine kinetics suggest that neither determines within-session operant responding. *Psychopharmacology, 183*(2), 201-209. doi:10.1007/s00213-005-0180-7

Cunningham, P., & Shahan, T. (2019). Rats engage in suboptimal choice when the delay to food is sufficiently long. *Journal of experimental psychology. Animal learning and cognition*.

Cunningham, P. J., & Shahan, T. A. (2018). Suboptimal choice, reward-predictive signals, and temporal information. *Journal of Experimental Psychology: Animal Learning and Cognition, 44*(1), 1.

D'Souza, M. S., & Duvauchelle, C. L. (2008). Certain or uncertain cocaine expectations influence accumbens dopamine responses toself-administered cocaine and non-rewarded operant behavior. *European Neuropsychopharmacology, 18*(9), 628-638.

Daniel, R., & Pollmann, S. (2014). A universal role of the ventral striatum in reward-based learning: evidence from human studies. *Neurobiology of learning and memory, 114*, 90-100.

Dart , R. C., Surratt , H. L., Cicero , T. J., Parrino , M. W., Severtson , S. G., Bucher-Bartelson , B., & Green , J. L. (2015). Trends in Opioid Analgesic Abuse and Mortality in the United States. *New England Journal of Medicine, 372*(3), 241-248. doi:10.1056/NEJMsa1406143

David, S. P., Munafò, M. R., Johansen-Berg, H., Smith, S. M., Rogers, R. D., Matthews, P. M., & Walton, R. T. (2005). Ventral striatum/nucleus accumbens activation to smoking-related pictorial cues in smokers and nonsmokers: a functional magnetic resonance imaging study. *Biological Psychiatry, 58*(6), 488-494.

Davis, D. R., Kurti, A. N., Skelly, J. M., Redner, R., White, T. J., & Higgins, S. T. (2016). A review of the literature on contingency management in the treatment of substance use disorders, 2009-2014. *Preventive medicine, 92*, 36-46. doi:10.1016/j.ypmed.2016.08.008

Davison, M., & McCarthy, D. (1988). *The matching law: A research review*: Lawrence Erlbaum Associates Inc.

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII, 23*(1).

Day, J. J., Jones, J. L., Wightman, R. M., & Carelli, R. M. (2010). Phasic nucleus accumbens dopamine release encodes effort-and delay-related costs. *Biological Psychiatry, 68*(3), 306-309.

Deroche-Gamonet, V., Belin, D., & Piazza, P. V. (2004). Evidence for addiction-like behavior in the rat. *Science, 305*(5686), 1014-1017.

Dickinson, A., Wood, N., & Smith, J. W. (2002). Alcohol seeking by rats: action or habit? *The Quarterly Journal of Experimental Psychology: Section B, 55*(4), 331-348.

Edwards, S., & Koob, G. F. (2013). Escalation of drug self-administration as a hallmark of persistent addiction liability. *Behavioural Pharmacology, 24*.

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation. *Brain research reviews, 62*(2), 233-244.

Estle, S. J., Green, L., Myerson, J., & Holt, D. D. (2006). Differential effects of amount on temporal and probability discounting of gains and losses. *Memory & Cognition, 34*(4), 914-928.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience, 8*(11), 1481.

Everitt, B. J., & Robbins, T. W. (2016). Drug addiction: updating actions to habits to compulsions ten years on. *Annual review of psychology, 67*, 23-50.

Fenno, L., Yizhar, O., & Deisseroth, K. (2011). The development and application of optogenetics. *Annual review of neuroscience, 34*.

Fiallos, A. M., Bricault, S. J., Cai, L. X., Worku, H. A., Colonnese, M. T., Westmeyer, G. G., & Jasanoff, A. (2017). Reward magnitude tracking by neural populations in ventral striatum. *NeuroImage, 146*, 1003-1015.

Fields, H. L., & Margolis, E. B. (2015). Understanding opioid reward. *Trends in Neurosciences, 38*(4), 217-225.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. *Science, 299*(5614), 1898-1902. doi:10.1126/science.1077349

Foster, T. M., Kinloch, J., & Poling, A. (2011). The effects of session length on demand functions generated using FR schedules. *Journal of the Experimental Analysis of Behavior, 95*(3), 289-304.

Francois, J., Conway, M. W., Lowry, J. P., Tricklebank, M. D., & Gilmour, G. (2012). Changes in reward-related signals in the rat nucleus accumbens measured by in vivo oxygen amperometry are consistent with fMRI BOLD responses in man. *NeuroImage, 60*(4), 2169-2181.

Francois, J., Huxter, J., Conway, M. W., Lowry, J. P., Tricklebank, M. D., & Gilmour, G. (2014). Differential contributions of infralimbic prefrontal cortex and nucleus accumbens during reward-based learning and extinction. *Journal of Neuroscience, 34*(2), 596-607.

Freeman, K. B., & Woolverton, W. L. (2011). Self-administration of cocaine and remifentanil by monkeys: choice between single drugs and mixtures. *Psychopharmacology, 215*(2), 281-290.

Galea, S., & Vlahov, D. (2002). Social determinants and the health of drug users: socioeconomic status, homelessness, and incarceration. *Public health reports, 117*(Suppl 1), S135.

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 37*(7), 1297-1310.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66*(4), 585-595.

Glass, P., Hardman, D., Kamiyama, Y., Quill, T. J., Marton, G., Donn, K., . . . Hermann, D. (1993). Preliminary pharmacokinetics and pharmacodynamics of an ultra-short-acting opioid: remifentanil (GI87084B). *Anesthesia and Analgesia, 77*(5), 1031-1040.

Glimcher, P. W. (2011). *Foundations of neuroeconomic analysis*: OUP USA.

Glimcher, P. W., Camerer, C. F., Fehr, E., & Poldrack, R. A. (2013). *Neuroeconomics: Decision making and the brain*: Academic Press.

Glimcher, P. W., Dorris, M. C., & Bayer, H. M. (2005). Physiological utility theory and the neuroeconomics of choice. *Games and economic behavior, 52*(2), 213-256.

Goeders, J. E., Murnane, K. S., Banks, M. L., & Fantegrossi, W. E. (2009). Escalation of food-maintained responding and sensitivity to the locomotor stimulant effects of cocaine in mice. *Pharmacology Biochemistry and Behavior, 93*(1), 67-74.

Goldberg, S. R. (1973). Comparable behavior maintained under fixed-ratio and second-order schedules of food presentation, cocaine injection or d-amphetamine injection in the squirrel monkey. *Journal of Pharmacology and Experimental Therapeutics, 186*(1), 18-30.

Goldstein, B. L., Barnett, B. R., Vasquez, G., Tobia, S. C., Kashtelyan, V., Burton, A. C., . . . Roesch, M. R. (2012). Ventral striatum encodes past and predicted value independent of motor contingencies. *Journal of Neuroscience, 32*(6), 2027-2036.

Grecksch, G., Bartzsch, K., Widera, A., Becker, A., Höllt, V., & Koch, T. (2006). Development of tolerance and sensitization to different opioid agonists in rats. *Psychopharmacology, 186*(2), 177-184.

Green, L., & Freed, D. E. (1993). The substitutability of reinforcers. *Journal of the Experimental Analysis of Behavior, 60*(1), 141-158.

Green, L., & Rachlin, H. (1991). Economic substitutability of electrical brain stimulation, food, and water. *Journal of the Experimental Analysis of Behavior, 55*(2), 133-143.

Griffiths, R. R., Bradford, L. D., & Brady, J. V. (1979). Progressive ratio and fixed ratio schedules of cocaine-maintained responding in baboons. *Psychopharmacology, 65*(2), 125-136.

Griffiths, R. R., Findley, J. D., Brady, J. V., Dolan-Gutcher, K., & Robinson, W. W. (1975). Comparison of progressive-ratio performance maintained by cocaine, methylphenidate and secobarbital. *Psychopharmacologia, 43*(1), 81-83.

Griffiths, R. R., Rush, C. R., & Puhala, K. A. (1996). Validation of the multiple-choice procedure for investigating drug reinforcement in humans. *Experimental and Clinical Psychopharmacology, 4*(1), 97.

Groman, S. M., Rich, K. M., Smith, N. J., Lee, D., & Taylor, J. R. (2017). Chronic Exposure to Methamphetamine Disrupts Reinforcement-Based Decision-Making in Rats. *Neuropsychopharmacology*. doi:10.1038/npp.2017.159

Guillem, K., Brenot, V., Durand, A., & Ahmed, S. H. (2018). Neuronal representation of individual heroin choices in the orbitofrontal cortex. *Addiction Biology, 23*(3), 880-888.

Haidar, S. H., Moreton, J. E., Liang, Z., Hoke, J. F., Muir, K. T., & EddingtonX, N. D. (1997). Evaluating a possible pharmacokinetic interaction between remifentanil and esmolol in the rat. *Journal of pharmaceutical sciences, 86*(11), 1278-1282.

Halbout, B., Liu, A. T., & Ostlund, S. B. (2016). A closer look at the effects of repeated cocaine exposure on adaptive decision-making under conditions that promote goal-directed control. *Frontiers in Psychiatry, 7*, 44.

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., . . . Berke, J. D. (2016). Mesolimbic dopamine signals the value of work. *Nature Neuroscience, 19*(1), 117.

Hart, A. S., Rutledge, R. B., Glimcher, P. W., & Phillips, P. E. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *Journal of Neuroscience, 34*(3), 698-704.

Hart, C. L., Haney, M., Foltin, R., & Fischman, M. (2000). Alternative reinforcers differentially modify cocaine self-administration by humans. *Behavioural Pharmacology, 11*(1), 87-91.

Heather, N. (2017). Is the concept of compulsion useful in the explanation or description of addictive behaviour and experience? *Addictive Behaviors Reports, 6*, 15-38.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4*(3), 267-272.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*(2), 243-266.

Herrnstein, R. J. (1974). Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior, 21*(1), 159-164.

Herrnstein, R. J., & Prelec, D. (1991). Melioration: A theory of distributed choice. *Journal of Economic Perspectives, 5*(3), 137-156.

Heyman, G. M. (2003). Consumption dependent changes in reward value: a framework for understanding addiction. In *Choice, behavioural economics and addiction* (pp. 95-127): Elsevier.

Heyman, G. M. (2013a). Addiction and choice: theory and new data. *Frontiers in Psychiatry, 4*, 31.

Heyman, G. M. (2013b). Addiction: an emergent consequence of elementary choice principles. *Inquiry, 56*(5), 428-445.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75*(4), 800-802.
Hodos, W. (1961). Progressive ratio as a measure of reward strength. *Science, 134*(3483), 943-944.
Hofford, R. S., Chow, J. J., Beckmann, J. S., & Bardo, M. T. (2017). Effects of environmental enrichment on self-administration of the short-acting opioid remifentanil in male rats. *Psychopharmacology, 234*(23-24), 3499-3506.
Hogarth, L. (2018). A Critical Review of Habit Theory of Drug Dependence. In *The Psychology of Habit* (pp. 325-341): Springer.
Holtyn, A. F., Koffarnus, M. N., DeFulio, A., Sigurdsson, S. O., Strain, E. C., Schwartz, R. P., & Silverman, K. (2014). Employment-based abstinence reinforcement promotes opiate and cocaine abstinence in out-of-treatment injection drug users. *Journal of Applied Behavior Analysis, 47*(4), 681-693.
Hooman Khademi, M., Farin Kamangar, M., Paul Brennan, M., & Reza Malekzadeh, M. (2016). Opioid therapy and its side effects: a review. *Archives of Iranian medicine, 19*(12), 870.
Hursh, S. R., & Silberberg, A. (2008). Economic demand and essential value. *Psychological Review, 115*(1), 186.
Hutto, C. W. J., & Crowder, W. F. (1997). Dosage choices of rats for morphine, for heroin, and between morphine and heroin. *Pharmacology Biochemistry and Behavior, 58*(1), 133-140.
Iigaya, K., Story, G. W., Kurth-Nelson, Z., Dolan, R. J., & Dayan, P. (2016). The modulation of savouring by prediction error and its effects on choice. *Elife, 5*, e13747.
Iversen, L., Iversen, S., Bloom, F. E., & Roth, R. H. (2008). *Introduction to neuropsychopharmacology*: Oxford University Press.
Jarmolowicz, D. P., Reed, D. D., DiGennaro Reed, F. D., & Bickel, W. K. (2016). The behavioral and neuroeconomics of reinforcer pathologies: Implications for managerial and health decision making. *Managerial and Decision Economics, 37*(4-5), 274-293.
Johanson, C. E., & Schuster, C. R. (1975). A choice procedure for drug reinforcers: cocaine and methylphenidate in the rhesus monkey. *Journal of Pharmacology and Experimental Therapeutics, 193*(2), 676-688.
Johnson, M. W., & Bickel, W. K. (2006). Replacing relative reinforcing efficacy with behavioral economic demand curves. *Journal of the Experimental Analysis of Behavior, 85*(1), 73-93.
Jones, J. D., & Comer, S. D. (2013). A review of human drug self-administration procedures. *Behavioural Pharmacology, 24*, 384.
Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience, 10*(12), 1625-1633. doi:10.1038/nn2007
Kacelnik, A. (2006). Meanings of Rationality. In M. Nudds & S. Hurley (Eds.), *Rational Animals?* (pp. 87-106). Oxford: Oxford University Press.
Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291. doi:10.2307/1914185

Kalivas, P., Volkow, N., & Seamans, J. (2005). Unmanageable motivation in addiction: a pathology in prefrontal-accumbens glutamate transmission. *Neuron, 45*(5), 647-650.
Kampman, K., & Jarvis, M. (2015). American Society of Addiction Medicine (ASAM) national practice guideline for the use of medications in the treatment of addiction involving opioid use. *Journal of addiction medicine, 9*(5), 358.
Katz, J. L. (1989). Drugs as reinforcers: Pharmacological and behavioural factors. *The neuropharmacological basis of reward*, 164-213.
Kearns, D. N., Kim, J. S., Tunstall, B. J., & Silberberg, A. (2017). Essential values of cocaine and non-drug alternatives predict the choice between them. *Addiction Biology, 22*(6), 1501-1514.
Kelleher, R., & Morse, W. (1964). *Escape behavior and punished behavior.* Paper presented at the Federation Proceedings.
Killeen, P. R. (1994). Mathematical principles of reinforcement. *Behavioral and Brain Sciences, 17*(1), 105-135.
Klingemann, H., Sobell, M. B., & Sobell, L. C. (2010). Continuities and changes in self-change research. *Addiction, 105*(9), 1510-1518.
Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *Journal of Neuroscience, 25*(19), 4806-4812.
Kobayashi, S., & Schultz, W. (2008). Influence of reward delays on responses of dopamine neurons. *Journal of Neuroscience, 28*(31), 7837-7846.
Koffarnus, M. N., & Woods, J. H. (2008). Quantification of drug choice with the generalized matching law in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 89*(2), 209-224.
Konova, A. B., Louie, K., & Glimcher, P. W. (2018). The computational form of craving is a selective multiplication of economic value. *Proceedings of the National Academy of Sciences, 115*(16), 4122-4127.
Kyonka, E. G. (2008). The matching law and effects of reinforcer rate and magnitude on choice in transition. *Behavioural Processes, 78*(2), 210-216.
Lak, A., Stauffer, W. R., & Schultz, W. (2014). Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proceedings of the National Academy of Sciences, 111*(6), 2343-2348.
Lamb, R., & Ginsburg, B. C. (2018). Addiction as a BAD, a behavioral allocation disorder. *Pharmacology Biochemistry and Behavior, 164*, 62-70.
Lamb, R., Maguire, D. R., Ginsburg, B. C., Pinkston, J. W., & France, C. P. (2016). Determinants of choice, and vulnerability and recovery in addiction. *Behavioural Processes, 127*, 35-42.
Landon, J., Davison, M., & Elliffe, D. (2003). Concurrent schedules: Reinforcer magnitude effects. *Journal of the Experimental Analysis of Behavior, 79*(3), 351-365.
Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 84*(3), 555-579.
Ledo, A., Lourenço, C. F., Laranjinha, J., Brett, C. M., Gerhardt, G. A., & Barbosa, R. M. (2017). Ceramic-based Multi-Site Platinum Microelectrode Arrays:

Morphological Characteristics and Electrochemical Performance for Extracellular Oxygen Measurements in Brain Tissue. *Analytical Chemistry*.
Lenoir, M., Cantin, L., Vanhille, N., Serre, F., & Ahmed, S. H. (2013). Extended heroin access increases heroin choices over a potent nondrug alternative. *Neuropsychopharmacology, 38*(7), 1209.
Lenoir, M., Serre, F., Cantin, L., & Ahmed, S. H. (2007). Intense sweetness surpasses cocaine reward. *PLoS ONE, 2*(8), e698.
Levine, A. I., & Bryson, E. O. (2010). Intranasal self-administration of remifentanil as the foray into opioid abuse by an anesthesia resident. *Anesthesia & Analgesia, 110*(2), 524-525.
Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology, 22*(6), 1027-1038.
Li, J., Schwarz, A. J., & Gilmour, G. (2015). Relating Translational Neuroimaging and Amperometric Endpoints: Utility for Neuropsychiatric Drug Discovery. In *Translational Neuropsychopharmacology* (pp. 397-421): Springer.
Li, Q., Wang, Y., Zhang, Y., Li, W., Yang, W., Zhu, J., . . . Qin, W. (2012). Craving correlates with mesolimbic responses to heroin-related cues in short-term abstinence from heroin: an event-related fMRI study. *Brain Research, 1469*, 63-72.
Liu, C. H., Greve, D. N., Dai, G., Marota, J. J., & Mandeville, J. B. (2007). Remifentanil administration reveals biphasic phMRI temporal responses in rat consistent with dynamic receptor regulation. *NeuroImage, 34*(3), 1042-1053.
Lloyd, C. (2013). The stigmatization of problem drug users: A narrative literature review. *Drugs: education, prevention and policy, 20*(2), 85-95.
Loree, A. M., Lundahl, L. H., & Ledgerwood, D. M. (2015). Impulsivity as a predictor of treatment outcome in substance use disorders: Review and synthesis. *Drug and alcohol review, 34*(2), 119-134.
Lowry, J. P., Griffin, K., McHugh, S. B., Lowe, A. S., Tricklebank, M., & Sibson, N. R. (2010). Real-time electrochemical monitoring of brain tissue oxygen: A surrogate for functional magnetic resonance imaging in rodents. *NeuroImage, 52*(2), 549-555. doi:http://dx.doi.org/10.1016/j.neuroimage.2010.04.187
Lubman, D. I., Yücel, M., & Pantelis, C. (2004). Addiction, a condition of compulsive behaviour? Neuroimaging and neuropsychological evidence of inhibitory dysregulation. *Addiction, 99*(12), 1491-1502.
Lucki, I. (1983). Rate-dependent effects of amphetamine on responding under random-interval schedules of reinforcement in the rat. *Pharmacology Biochemistry and Behavior, 18*(2), 195-201.
Lussier, J. P., Heil, S. H., Mongeon, J. A., Badger, G. J., & Higgins, S. T. (2006). A meta-analysis of voucher-based reinforcement therapy for substance use disorders. *Addiction, 101*(2), 192-203.
Mackillop, J., Amlung, M. T., Few, L. R., Ray, L. A., Sweet, L. H., & Munafo, M. R. (2011). Delayed reward discounting and addictive behavior: a meta-analysis. *Psychopharmacology, 216*, 305-321. doi:10.1007/s00213-011-2229-0
Madden, G. J., Bickel, W. K., & Jacobs, E. A. (2000). Three predictions of the economic concept of unit price in a choice context. *Journal of the Experimental Analysis of Behavior, 73*(1), 45-64.

Maguire, D. R., Gerak, L. R., & France, C. P. (2013). Delay discounting of food and remifentanil in rhesus monkeys. *Psychopharmacology, 229*(2), 323-330.
Marchant, N. J., Li, X., & Shaham, Y. (2013). Recent developments in animal models of drug relapse. *Current Opinion in Neurobiology, 23*(4), 675-683.
Marshall, A. T., & Kirkpatrick, K. (2013). The effects of the previous outcome on probabilistic choice in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 39*(1), 24-38.
Marshall, A. T., & Kirkpatrick, K. (2017). Reinforcement learning models of risky choice and the promotion of risk-taking by losses disguised as wins in rats. *Journal of Experimental Psychology: Animal Learning and Cognition, 43*(3), 262.
Marshall, A. T., Smith, A. P., & Kirkpatrick, K. (2014). Mechanisms of impulsive choice: I. Individual differences in interval timing and reward processing. *Journal of the Experimental Analysis of Behavior*. doi:10.1002/jeab.88
Mazur, J. E. (2006). Mathematical models and the experimental analysis of behavior. *Journal of the Experimental Analysis of Behavior, 85*(2), 275-291.
McCauley, R. N. (2007). Reduction: Models of cross-scientific relations and their implications for the psychology-neuroscience interface. In *Philosophy of psychology and cognitive science* (pp. 105-158): Elsevier.
McDevitt, M. A., Dunn, R. M., Spetch, M. L., & Ludvig, E. A. (2016). When good news leads to bad choices. *Journal of the Experimental Analysis of Behavior, 105*(1), 23-40. doi:10.1002/jeab.192
McDowell, J. (2005). On the classic and modern theories of matching. *Journal of the Experimental Analysis of Behavior, 84*(1), 111-127.
McDowell, J. (2013). On the theoretical and empirical status of the matching law and matching theory. *Psychological Bulletin, 139*(5), 1000.
McLean, A. P., & Blampied, N. M. (2001). Sensitivity to relative reinforcer rate in concurrent schedules: Independence from relative and absolute reinforcer duration. *Journal of the Experimental Analysis of Behavior, 75*(1), 25-42.
McMillan, D. (1969). Effects of d-amphetamine on performance under several parameters of multiple fixed-ratio, fixed-interval schedules. *Journal of Pharmacology and Experimental Therapeutics, 167*(1), 26-33.
Mello, N. K., & Negus, S. S. (1996). Preclinical evaluation of pharmacotherapies for treatment of cocaine and opioid abuse using drug self-administration procedures. *Neuropsychopharmacology, 14*(6), 375.
Mercer, M. E., & Holder, M. D. (1997). Food cravings, endogenous opioid peptides, and food intake: a review. *Appetite, 29*(3), 325-352.
Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117*(3), 363.
Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of neurophysiology, 72*(2), 1024-1027.
Moeller, S. J., & Stoops, W. W. (2015). Cocaine choice procedures in animals, humans, and treatment-seekers: Can we bridge the divide? *Pharmacology Biochemistry and Behavior, 138*, 133-141.
Molet, M., Miller, H., Laude, J., Kirk, C., Manning, B., & Zentall, T. (2012). Decision making by humans in a behavioral task: Do humans, like pigeons, show

suboptimal choice? *Learning & Behavior, 40*(4), 439-447. doi:10.3758/s13420-012-0065-7
Morgan, C. T., Stellar, E., & Johnson, O. (1943). Food-deprivation and hoarding in rats. *Journal of Comparative Psychology, 35*(3), 275.
Murphy, J. G., MacKillop, J., Skidmore, J. R., & Pederson, A. A. (2009). Reliability and validity of a demand curve measure of alcohol reinforcement. *Experimental and Clinical Psychopharmacology, 17*(6), 396.
Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology, 47*(1), 90-100.
N Kearns, D., A Gomez-Serrano, M., & J Tunstall, B. (2011). A review of preclinical research demonstrating that drug and Non-Drug ReinforcersDifferentially affect behavior. *Current drug abuse reviews, 4*(4), 261-269.
National Drug Intelligence Center Product, U. S. D. o. J. (2011). *National Drug Threat Assessment. Product No. 2011-Q0317-001*. Retrieved from
Negus, S. S. (2005). Effects of punishment on choice between cocaine and food in rhesus monkeys. *Psychopharmacology, 181*(2), 244-252.
Negus, S. S. (2006). Choice between heroin and food in nondependent and heroin-dependent rhesus monkeys: effects of naloxone, buprenorphine, and methadone. *Journal of Pharmacology and Experimental Therapeutics, 317*(2), 711-723.
Negus, S. S., & Rice, K. C. (2009). Mechanisms of withdrawal-associated increases in heroin self-administration: pharmacologic modulation of heroin vs food choice in heroin-dependent rhesus monkeys. *Neuropsychopharmacology, 34*(4), 899.
Neuringer, A. J. (1967). Effects of reinforcement magnitude on choice and rate of responding. *Journal of the Experimental Analysis of Behavior, 10*(5), 417-424.
O'Doherty, J. P., Buchanan, T. W., Seymour, B., & Dolan, R. J. (2006). Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron, 49*(1), 157-166.
O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences, 1104*(1), 35-53.
Odum, A. L. (2011). Delay discounting: I'm a K, you're a K. *Journal of the Experimental Analysis of Behavior, 96*(3), 427-439. doi:10.1901/jeab.2011.96-423
Ojeda, A., Murphy, R. A., & Kacelnik, A. (2018). Paradoxical choice in rats: Subjective valuation and mechanism of choice. *Behavioural Processes, 152*, 73-80.
Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441*(7090), 223.
Pareto, V. (1906/1971). *General notion of economic equilibrium [translated by A.S. Schwier, 1971, Manuale di Economia Politica, Societa Editrice Libari,Milan.* New York: Macmillan.
Patros, C. H., Alderson, R. M., Kasper, L. J., Tarle, S. J., Lea, S. E., & Hudec, K. L. (2016). Choice-impulsivity in children and adolescents with attention-deficit/hyperactivity disorder (ADHD): A meta-analytic review. *Clinical Psychology Review, 43*, 162-174.
Paxinos, G., & Watson, C. (2006). *The rat brain in stereotaxic coordinates: hard cover edition*: Elsevier.

Pelloux, Y., Murray, J. E., & Everitt, B. J. (2015). Differential vulnerability to the punishment of cocaine related behaviours: effects of locus of punishment, cocaine taking history and alternative reinforcer availability. *Psychopharmacology, 232*(1), 125-134.

Phillips, P. E., Stuber, G. D., Heien, M. L., Wightman, R. M., & Carelli, R. M. (2003). Subsecond dopamine release promotes cocaine seeking. *Nature, 422*(6932), 614.

Picker, M., & Poling, A. (1982). Choice as a dependent measure in autoshaping: Sensitivity to frequency and duration of food presentation. *Journal of the Experimental Analysis of Behavior, 37*(3), 393-406. doi:10.1901/jeab.1982.37-393

Pierce, R. C., & Kumaresan, V. (2006). The mesolimbic dopamine system: the final common pathway for the reinforcing effect of drugs of abuse? *Neuroscience & Biobehavioral Reviews, 30*(2), 215-238.

Pinheiro, J., Bates, D., DebRoy, S., & Team, R. C. (2016). *nlme: Linear and nonlinear mixed effects models*. *R package version 3.1-128*.

Powell, R. A., Honey, P. L., & Symbaluk, D. G. (2016). *Introduction to learning and behavior*: Cengage Learning.

Rachlin, H. (1971). On the tautology of the matching law. *Journal of the Experimental Analysis of Behavior, 15*(2), 249-251.

Rachlin, H. (1997). Four teleological theories of addiction. *Psychonomic Bulletin & Review, 4*(4), 462-473.

Rachlin, H. (2000). The lonely addict. *Reframing health behavior change with behavioral economics*, 145-164.

Rachlin, H., Battalio, R., Kagel, J., & Green, L. (1981). Maximization theory in behavioral psychology. *Behavioral and Brain Sciences, 4*(3), 371-388.

Rachlin, H., Green, L., Kagel, J. H., & Battalio, R. C. (1976). Economic demand theory and psychological studies of choice. In *Psychology of Learning and Motivation* (Vol. 10, pp. 129-154): Elsevier.

Rasmussen, E. B., Lawyer, S. R., & Reilly, W. (2010). Percent body fat is related to delay and probability discounting for food in humans. *Behavioural Processes, 83*(1), 23-30. doi:http://dx.doi.org/10.1016/j.beproc.2009.09.001

Regier, P. S., & Redish, A. D. (2015). Contingency management and deliberative decision-making processes. *Frontiers in Psychiatry, 6*, 76.

Renda, C. R., & Madden, G. J. (2016). Impulsive choice and pre-exposure to delays: III. Four-month test-retest outcomes in male wistar rats. *Behavioural Processes, 126*, 108-112.

Renda, C. R., Stein, J. S., & Madden, G. J. (2015). Working-memory training: Effects on delay discounting in male long evans rats. *Journal of the Experimental Analysis of Behavior, 103*(1), 50-61. doi:10.1002/jeab.115

Rescorla, R. A. (2003). Contemporary study of Pavlovian conditioning. *The Spanish journal of psychology, 6*(02), 185-195.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory, 2*, 64-99.

Reynolds, R. W. (1958). The relationship between stimulation voltage and rate of hypothalamic self-stimulation in the rat. *Journal of Comparative and Physiological Psychology, 51*(2), 193.
Richardson, N. R., & Roberts, D. C. (1996). Progressive ratio schedules in drug self-administration studies in rats: a method to evaluate reinforcing efficacy. *Journal of neuroscience methods, 66*(1), 1-11.
Risinger, R. C., Salmeron, B. J., Ross, T. J., Amen, S. L., Sanfilipo, M., Hoffmann, R. G., . . . Stein, E. A. (2005). Neural correlates of high and craving during cocaine self-administration using BOLD fMRI. *NeuroImage, 26*(4), 1097-1108.
Robins, L. N. (1993). Vietnam veterans' rapid recovery from heroin addiction: A fluke or normal expectation? *Addiction, 88*(8), 1041-1054.
Robinson, T. E., & Berridge, K. C. (2008). Review. The incentive sensitization theory of addiction: some current issues. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 363*(1507), 3137-3146. doi:10.1098/rstb.2008.0093
Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience, 10*(12), 1615.
Rush, C. R., & Stoops, W. W. (2012). Agonist replacement therapy for cocaine dependence: a translational review. *Future medicinal chemistry, 4*(2), 245-265.
Rutherford, E. C., Pomerleau, F., Huettl, P., Strömberg, I., & Gerhardt, G. A. (2007). Chronic second-by-second measures of l-glutamate in the central nervous system of freely moving rats. *Journal of neurochemistry, 102*(3), 712-722.
Rutledge, R. B., Lazzaro, S. C., Lau, B., Myers, C. E., Gluck, M. A., & Glimcher, P. W. (2009). Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *Journal of Neuroscience, 29*(48), 15104-15114.
Saddoris, M. P., Cacciapaglia, F., Wightman, R. M., & Carelli, R. M. (2015a). Differential dopamine release dynamics in the nucleus accumbens core and shell reveal complementary signals for error prediction and incentive motivation. *Journal of Neuroscience, 35*(33), 11572-11582.
Saddoris, M. P., Sugam, J. A., Stuber, G. D., Witten, I. B., Deisseroth, K., & Carelli, R. M. (2015b). Mesolimbic dopamine dynamically tracks, and is causally linked to, discrete aspects of value-based decision making. *Biological Psychiatry, 77*(10), 903-911.
Saddoris, M. P., Wang, X., Sugam, J. A., & Carelli, R. M. (2016). Cocaine Self-Administration Experience Induces Pathological Phasic Accumbens Dopamine Signals and Abnormal Incentive Behaviors in Drug-Abstinent Rats. *The Journal of Neuroscience, 36*(1), 235-250. doi:10.1523/jneurosci.3468-15.2016
Salgado, S., & Kaplitt, M. G. (2015). The nucleus accumbens: a comprehensive review. *Stereotactic and functional neurosurgery, 93*(2), 75-93.
SAMHSA. (2017). Key substance use and mental health indicators in the United States: Results from the 2016 National Survey on Drug Use and Health. Retrieved from http://www.samshsa.gov/data

Sanchis-Segura, C., & Spanagel, R. (2006). Behavioural assessment of drug reinforcement and addictive features in rodents: an overview. *Addiction Biology, 11*(1), 2-38.
Schultz, W. (2010). Dopamine signals for reward value and risk: basic and recent data. *Behavioral and Brain Functions, 6*(1), 24.
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.
Schultz, W., Apicella, P., Scarnati, E., & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *Journal of Neuroscience, 12*(12), 4595-4610.
Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593-1599.
Schwartz, L. P., Kim, J. S., Silberberg, A., & Kearns, D. N. (2017). Heroin and saccharin demand and preference in rats. *Drug and Alcohol Dependence, 178*, 87-93.
Schwartz, L. P., Roma, P. G., Henningfield, J. E., Hursh, S. R., Cone, E. J., Buchhalter, A. R., . . . Schnoll, S. H. (2019). Behavioral economic demand metrics for abuse deterrent and abuse potential quantification. *Drug and Alcohol Dependence*.
Scofield, M., Heinsbroek, J., Gipson, C., Kupchik, Y., Spencer, S., Smith, A., . . . Kalivas, P. (2016). The nucleus accumbens: mechanisms of addiction across drug classes reflect the importance of glutamate homeostasis. *Pharmacological reviews, 68*(3), 816-871.
Sheth, S. A., Nemoto, M., Guiou, M., Walker, M., Pouratian, N., & Toga, A. W. (2004). Linear and nonlinear relationships between neuronal activity, oxygen metabolism, and hemodynamic responses. *Neuron, 42*(2), 347-355.
Siegel, S. (2005). Drug tolerance, drug addiction, and drug anticipation. *Current Directions in Psychological Science, 14*(6), 296-300.
Skinner, B. (1932). Drive and reflex strength. *The Journal of General Psychology, 6*(1), 22-37.
Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.
Smethells, J. R., Harris, A. C., Burroughs, D., Hursh, S. R., & LeSage, M. G. (2018). Substitutability of nicotine alone and an electronic cigarette liquid using a concurrent choice assay in rats: A behavioral economic analysis. *Drug and Alcohol Dependence, 185*, 58-66.
Smith, A. P., Beckmann, J. S., & Zentall, T. R. (2017a). Gambling-like behavior in pigeons: 'jackpot' signals promote maladaptive risky choice. *Scientific reports*.
Smith, A. P., Hofford, R. S., Zentall, T. R., & Beckmann, J. S. (2018). The role of 'jackpot'stimuli in maladaptive decision-making: dissociable effects of D1/D2 receptor agonists and antagonists. *Psychopharmacology*, 1-11.
Smith, A. P., Marshall, A. T., & Kirkpatrick, K. (2015). Mechanisms of impulsive choice: II. Time-based interventions to improve self-control. *Behavioural Processes, 112*(0), 29-42. doi:http://dx.doi.org/10.1016/j.beproc.2014.10.010
Smith, M. A., Greene-Naples, J. L., Lyle, M. A., Iordanou, J. C., & Felder, J. N. (2009). The effects of repeated opioid administration on locomotor activity: I. Opposing

actions of μ and κ receptors. *Journal of Pharmacology and Experimental Therapeutics, 330*(2), 468-475.
Smith, T. R., Beran, M. J., & Young, M. E. (2017b). Gambling in rhesus macaques (Macaca mulatta): The effect of cues signaling risky choice outcomes. *Learning & Behavior, 45*(3), 288-299. doi:10.3758/s13420-017-0270-5
Smith, T. T., Rupprecht, L. E., Sved, A. F., & Donny, E. C. (2016). Characterizing the relationship between increases in the cost of nicotine and decreases in nicotine content in adult male rats: implications for tobacco regulation. *Psychopharmacology, 233*(23-24), 3953-3964.
Spiga, R., Maxwell, R. S., Meisch, R. A., & Grabowski, J. (2005). Human methadone self-administration and the generalized matching law. *The Psychological Record, 55*(4), 525-538.
Spragg, S. D. S. (1940). Morphine addiction in chimpanzees. *Comparative Psychology Monographs*.
Stagner, J., & Zentall, T. (2010). Suboptimal choice behavior by pigeons. *Psychonomic Bulletin & Review, 17*(3), 412-416. doi:10.3758/PBR.17.3.412
Stephens, D. W. (2008). Decision ecology: foraging and the ecology of animal decision making. *Cognitive, affective, & behavioral neuroscience, 8*(4), 475-484.
Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*(3), 153-181. doi:http://dx.doi.org/10.1037/h0046162
Stopper, C. M., Maric, T., Montes, D. R., Wiedman, C. R., & Floresco, S. B. (2014). Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron, 84*(1), 177-189.
Sugam, J. A., Day, J. J., Wightman, R. M., & Carelli, R. M. (2012). Phasic nucleus accumbens dopamine encodes risk-based decision-making behavior. *Biological Psychiatry, 71*(3), 199-205.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1): MIT press Cambridge.
Symonds, M. R., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology, 65*(1), 13-21.
Tatham, T. A., & Zurn, K. R. (1989). The MED-PC experimental apparatus programming system. *Behavior Research Methods, Instruments & Computers, 21*(2), 294-302.
Thompson, T., Honor, J., Verchota, S., & Cleary, J. (1984). Interval and ratio reinforcement contingencies as determinants of methadone's effects. *Pharmacology Biochemistry and Behavior, 21*(5), 743-747.
Trescot, A. M., Datta, S., Lee, M., & Hansen, H. (2008). Opioid pharmacology. *Pain physician, 11*(2 Suppl), S133-153.
Trujano, R. E., López, P., Rojas-Leguizamón, M., & Orduña, V. (2016). Optimal behavior by rats in a choice task is associated to a persistent conditioned inhibition effect. *Behavioural Processes, 130*, 65-70.
Trujano, R. E., & Orduna, V. (2015). Rats are optimal in a choice task in which pigeons are not. *Behavioural Processes, 119*, 22-27.
Tunstall, B. J., & Kearns, D. N. (2014). Cocaine can generate a stronger conditioned reinforcer than food despite being a weaker primary reinforcer. *Addiction Biology*, n/a-n/a. doi:10.1111/adb.12195

Urban, D. J., & Roth, B. L. (2015). DREADDs (designer receptors exclusively activated by designer drugs): chemogenetic tools with therapeutic utility. *Annu Rev Pharmacol Toxicol, 55*, 399-417.
Urcuioli, P. J. (2005). Behavioral and associative effects of differential outcomes in discrimination learning. *Animal Learning & Behavior, 33*(1), 1-21.
Vandaele, Y., Cantin, L., Serre, F., Vouillac-Mendoza, C., & Ahmed, S. H. (2016). Choosing under the influence: a drug-specific mechanism by which the setting controls drug choices in rats. *Neuropsychopharmacology, 41*(2), 646.
Vandaele, Y., & Janak, P. H. (2018). Defining the place of habit in substance use disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 87*, 22-32.
Vanderschuren, L. J., & Everitt, B. J. (2004). Drug seeking becomes compulsive after prolonged cocaine self-administration. *Science, 305*(5686), 1017-1019.
Vasconcelos, M., Monteiro, T., & Kacelnik, A. (2015). Irrational choice and the value of information. *Scientific reports, 5*, 13874. doi:10.1038/srep13874
Venniro, M., Zhang, M., Caprioli, D., Hoots, J. K., Golden, S. A., Heins, C., . . . Shaham, Y. (2018). Volitional social interaction prevents drug addiction in rat models. *Nature Neuroscience, 21*(11), 1520.
Venniro, M., Zhang, M., Shaham, Y., & Caprioli, D. (2017). Incubation of methamphetamine but not heroin craving after voluntary abstinence in male and female rats. *Neuropsychopharmacology, 42*(5), 1126.
Volkow, N. D., Koob, G. F., & McLellan, A. T. (2016). Neurobiologic advances from the brain disease model of addiction. *New England Journal of Medicine, 374*(4), 363-371.
Volkow, N. D., Wang, G. J., Fowler, J. S., & Tomasi, D. (2012). Addiction circuitry in the human brain. *Annu Rev Pharmacol Toxicol, 52*, 321-336. doi:10.1146/annurev-pharmtox-010611-134625
Vowles, K. E., McEntee, M. L., Julnes, P. S., Frohe, T., Ney, J. P., & van der Goes, D. N. (2015). Rates of opioid misuse, abuse, and addiction in chronic pain: a systematic review and data synthesis. *Pain, 156*(4), 569-576.
Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192-196.
Wallis, J. D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience, 15*(1), 13.
Wearden, J., & Burgess, I. (1982). Matching since baum (1979). *Journal of the Experimental Analysis of Behavior, 38*(3), 339-348.
Weeks, J. R. (1962). Experimental morphine addiction: method for automatic intravenous injections in unrestrained rats. *Science, 138*(3537), 143-144.
Williams, C. T., & Latkin, C. A. (2007). Neighborhood socioeconomic status, personal network attributes, and use of heroin and cocaine. *American journal of preventive medicine, 32*(6), S203-S210.
Williams, J. T., Christie, M. J., & Manzoni, O. (2001). Cellular and synaptic adaptations mediating opioid dependence. *Physiological Reviews, 81*(1), 299-343.
Wise, R. A., & Koob, G. F. (2014). The development and maintenance of drug addiction. *Neuropsychopharmacology, 39*(2), 254.

Witkin, J. M. (1993). Blockade of the locomotor stimulant effects of cocaine and methamphetamine by glutamate antagonists. *Life sciences, 53*(24), PL405-PL410.
Woolverton, W. L., & Rowlett, J. K. (1998). Choice maintained by cocaine or food in monkeys: effects of varying probability of reinforcement. *Psychopharmacology, 138*(1), 102-106.
Wunderlich, K., Rangel, A., & O'Doherty, J. P. (2009). Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences, 106*(40), 17199-17204.
Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D. F., & Büchel, C. (2006). Dissociable systems for gain-and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience, 26*(37), 9530-9537.
Yates, J. R., Breitenstein, K. A., Gunkel, B. T., Hughes, M. N., Johnson, A. B., Rogers, K. K., & Shape, S. M. (2016). Effects of NMDA receptor antagonists on probability discounting depend on the order of probability presentation. *Pharmacology Biochemistry and Behavior, 150–151*, 31-38. doi:http://dx.doi.org/10.1016/j.pbb.2016.09.004
Young, M. E. (2017). Discounting: A Practical Guide to Multilevel Analysis of Indifference Data. *Journal of the Experimental Analysis of Behavior*.
Young, M. E., Clark, M., Goffus, A., & Hoane, M. R. (2009). Mixed effects modeling of Morris water maze data: Advantages and cautionary notes. *Learning and Motivation, 40*(2), 160-177.
Zentall, T. R. (2016). Resolving the paradox of suboptimal choice. *Journal of Experimental Psychology: Animal Learning and Cognition, 42*(1), 1.
Zentall, T. R., & Stagner, J. (2011). Maladaptive choice behaviour by pigeons: an animal analogue and possible mechanism for gambling (sub-optimal human decision-making behaviour). *Proceedings of the Royal Society B: Biological Sciences, 278*(1709), 1203-1208.

CURRICULUM VITAE

Aaron Paul Smith

Education

2016-2019

University of Kentucky, Lexington, KY

Graduate Student, M.S.

Advisor: Dr. Joshua S. Beckmann

Expected Graduation: August, 2019

2013 – 2016

University of Kentucky, Lexington, KY

Graduate Student

Advisor: Dr. Thomas R. Zentall

2009 - 2013

Kansas State University, Manhattan, KS

Bachelor of Arts in Psychology with Magna Cum Laude

Minors in German, Leadership Studies, and History

Teaching Experience

2019

Brain and Behavior – University of Kentucky, Department of Psychology

Guest Lecturer on Neuroanatomy and Action Potential Propagation

2017

Experimental Psychology of Learning – Transylvania University, Department of Psychology

Instructor on Record

Learning and Memory - University of Kentucky, Department of Psychology

Instructor on Record

2013 – 2015

Graduate Teaching Assistant – University of Kentucky, Department of Psychology

Introductory Psychology, Learning & Memory (online), Learning, Evolutionary Psychology, Cognitive Psychology

Research Experience

2015 – Present

Graduate Research Assistant – University of Kentucky Department of Psychology

Dr. Joshua Beckmann: Behavioral Neuroscience Laboratory

2013 – 2016

Graduate Research Assistant – University of Kentucky Department of Psychology

Dr. Thomas Zentall: Comparative Cognition Laboratory

2011 - 2013

Undergraduate Research Assistant – Kansas State Department of Psychological Sciences

Dr. Kimberly Kirkpatrick: Reward, Timing, and Decision Laboratory

Publications

Zentall, T. R., **Smith, A. P.,** & Beckmann, J. S. (2019). Differences in rats and pigeons suboptimal choice may depend on where those stimuli are in their behavioral system. *Behavioral Processes, 159*, 37-41. doi: 10.1016/j.beproc.2018.11.012

Smith, A.P., Zentall, T. R., & Kacelnik A. (2018) Midsession reversal task with pigeons: Parallel processing of alternatives explains choices. *Journal of Experimental Psychology: Animal Learning and Cognition, 44*(3), 272-279. doi: 10.1037/xan0000180

Smith, A. P., Hofford, R. S., Zentall, T. R., & Beckmann, J. S. (2018). The role of 'jackpot' stimuli in maladaptive decision-making: dissociable effects of D1/D2 receptor agonists and antagonists. *Psychopharmacology*, *5*, 1427-1437.

Smith, A. P., Beckmann, J. S., & Zentall, T. R. (2017). Gambling-like behavior in pigeons: 'jackpot' signals promote maladaptive risky choice. *Scientific Reports, 7(1)*, 6625. doi: 10.1038/s41598-017-06641-x

Smith, A. P., Beckmann, J. S., & Zentall, T. R. (2017). Mechanisms of midsession reversal accuracy: Memory for preceding events and timing. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*(1), 62.

Chow, J. J., **Smith, A. P.,** Wilson, A. G., Zentall, T. R., & Beckmann, J. S. (2017). Suboptimal choice in rats: Incentive salience attribution promotes maladaptive decision-making. *Behavioural Brain Research, 320*, 244-254. doi: http://dx.doi.org/10.1016/j.bbr.2016.12.013

Smith, A. P., Bailey, A. R., Chow, J. J., Beckmann, J. S., & Zentall, T. R. (2016). Suboptimal choice in pigeons: Stimulus value predicts choice over frequencies. *Plos one*, *11*(7), e0159336.

Smith, A. P., Peterson, J. R., & Kirkpatrick, K. (2016). Reward Contrast Effects on Impulsive Choice and Timing in Rats. *Timing & Time Perception*. doi: 10.1163/22134468-00002059

Smith, A. P., & Zentall, T. R. (2016). Suboptimal Choice in Pigeons: Choice Is Primarily Based on the Value of the Conditioned Reinforcer Rather Than Overall Reinforcement Rate. *Journal of Experimental Psychology: Animal Learning and Cognition, 42(2)*, 212-220. doi: http://dx.doi.org/10.1037/xan0000092

Zentall, T. R. & **Smith, A. P.** (2016). Delayed Matching-to-Sample: A Tool to Assess Memory and Other Cognitive Processes. *Behavioral Processes, 123*, 26-42. doi: 10.1016/j.beproc.2015.07.002

Smith, A. P., Pattison, K. F., & Zentall, T. R. (2016). Rats' midsession reversal performance: the nature of the response. *Learning & Behavior*, 1-10. doi: 10.3758/s13420-015-0189-7

Zentall, T. R., Laude, J. R., Stagner, J. P., & **Smith, A. P.** (2015). Suboptimal choice by pigeons: Evidence that the value of the conditioned reinforcer rather than its frequency determines choice. *The Psychological Record*, *65*(2), 223-229.

Kirkpatrick, K., Marshall, A. T., & **Smith, A. P.** (2015). Mechanisms of Individual Differences in Impulsive and Risky Choice in Rats. *Comparative Cognition and Behavior Reviews, 10*, 45-72. doi: 10.3819/ccbr.2015.100003

Smith, A. P., Marshall, A. T., & Kirkpatrick, K. (2015). Mechanisms of impulsive choice: II. Time-based interventions to improve self-control. *Behavioural Processes, 112*(0), 29-42. doi: http://dx.doi.org/10.1016/j.beproc.2014.10.010

Zentall, T. R., & **Smith, A. P.** (2014). Risk should be objectively defined: comment on Pelé and Sueur. *Animal cognition*, *17*(6), 1433-1436.

Kirkpatrick, K., Marshall, A. T., **Smith, A. P.**, Koci, J., & Park, Y. (2014). Individual differences in impulsive and risky choice: Effects of environmental rearing conditions. *Behavioural Brain Research*, *269*, 115-127. doi: http://dx.doi.org/10.1016/j.bbr.2014.04.024

Marshall, A. T., **Smith, A. P.**, & Kirkpatrick, K. (2014). Mechanisms of impulsive choice: I. Individual differences in interval timing and reward processing. *Journal of the Experimental Analysis of Behavior*. doi: 10.1002/jeab.88

Scholarships and Awards

2017

T32 in Drug Abuse Research (DA016176)

2015

University of Kentucky, Department of Psychology Outstanding Graduate Student in Cognition, Learning, and Performance

2013

Theta Xi Academic Scholarship

2012

Doreen Shanteau Undergraduate Research Fellowship

Marching Pride Scholar

Pride Leader Scholarship

2011

Study Abroad Scholarship

2009

Kansas State University Leadership Scholarship