Psychopharmacology (2021) 238:2047–2057
https://doi.org/10.1007/s00213-021-05830-x

ORIGINAL INVESTIGATION

Quantifying value-based determinants of drug and non-drug decision dynamics

Aaron P. Smith[1] · Joshua S. Beckmann[2]

Received: 16 November 2020 / Accepted: 15 March 2021 / Published online: 10 April 2021

Abstract

Rationale A growing body of research suggests that substance use disorder (SUD) may be characterized as disorders of decision making. However, drug choice studies assessing drug-associated decision making often lack more complex and dynamic conditions that better approximate contexts outside the laboratory and may lead to incomplete conclusions regarding the nature of drug-associated value.

Objectives The current study assessed isomorphic (choice between identical food options) and allomorphic (choice between remifentanil [REMI] and food) choice across dynamically changing reward probabilities, magnitudes, and differentially reward-predictive stimuli in male rats to better understand determinants of drug value. Choice data were analyzed at aggregate and choice-by-choice levels using quantitative matching and reinforcement learning (RL) models, respectively.

Results Reductions in reward probability or magnitude independently reduced preferences for food and REMI commodities. Inclusion of reward-predictive cues significantly increased preference for food and REMI rewards. Model comparisons revealed that reward-predictive stimuli significantly altered the economic substitutability of food and REMI rewards at both levels of analysis. Furthermore, model comparisons supported the reformulation of reward value updating in RL models from independent terms to a shared, relative term, more akin to matching models.

Conclusions The results indicate that value-based quantitative choice models can accurately capture choice determinants within complex decision-making contexts and corroborate drug choice as a multidimensional valuation process. Collectively, the present study indicates commonalities in decision-making for drug and non-drug rewards, validates the use of economic-based SUD therapies (e.g., contingency management), and implicates the neurobehavioral processes underlying drug-associated decision-making as a potential avenue for future SUD treatment.

Keywords Reinforcement learning · Drug · Rat · Matching · Habit · Remifentanil · Opioid · Choice

Introduction

Growing evidence suggests substance use disorder (SUD), including opioid use disorder, consists, at least in part, of disordered decision-making mechanisms (Ahmed 2005; Beckmann et al. 2019; Davis et al. 2016; Heather 2017; Hogarth 2020; Hogarth and Field 2020; Moeller and Stoops 2015). Thus, this growing evidence implicates decision-making processes as potential therapeutic targets for SUD. Relatedly, to study the neurobehavioral mechanisms of decision-making, choice experiments have long been the standard protocol. Recent studies have attempted to improve the validity of choice experiments as models of real-world decision making via the implementation of dynamic decision-making procedures (Corrado et al. 2009). Within these dynamic contexts, the contingencies associated with each choice option change unpredictably and require the constant engagement of decision-making processes over a large number of choice opportunities. The rich datasets produced by dynamic procedures then afford formalization of decision-making processes through quantitative choice modeling. For instance, the relations between reward-associated brain signals to specific predictions of reinforcement learning (RL; Glimcher 2011) and matching models of choice (Sugrue et al. 2004) are now

✉ Joshua S. Beckmann
joshua.beckmann@uky.edu

[1] Cofrin Logan Center for Addiction Research and Treatment, University of Kansas, Lawrence, KS, USA

[2] Department of Psychology, University of Kentucky, Lexington, KY, USA

prominent relationships founded upon quantitative modeling of choice dynamics, highlighting the benefits of a modeling approach in choice studies. Importantly, these results provided evidence for the existence of a biological architecture that may construct choice preferences consistent with quantitative choice models that are based primarily on behavioral data.

Despite the heavy use of dynamic choice procedures coupled with formal quantitative modeling in decision sciences, these methodologies are largely absent in the SUD literature, including studies of drug choice. Yet, given the successful history of the approach in improving our understanding of decision-making, the application of such methodology may help inform mechanisms of drug preference while also improving the translational value of preclinical drug choice studies (Ahmed 2010; Banks et al. 2015). For instance, quantitative modeling of choice data can help index the effects of drug exposure to specific facets of decision-making. As an example, Eq. 1 shows a proportional form of generalized matching that describes choices between commodities varying in both probability (or rate) and amount (or magnitude).

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

In Eq. 1, choices for a commodity, such as an opioid drug reward like remifentanil (REMI; B_A), are determined by the ratio of REMI reward probability (R) and magnitude (M) relative to other available commodities, such as food (B_B). Each reinforcer dimension is then scaled with sensitivity parameters (S_R, S_M) that determine how quickly choices redistribute as the ratios change. Thus, by using models like Eq. 1, the effect of REMI on sensitivities to specific reinforcer dimensions can be isolated.

RL models can also help isolate specific facets of decision-making, but do so at a choice-by-choice resolution that constructs the decision process according to reward prediction errors as shown below (RPE; Glimcher 2011; Sutton and Barto 1998).

$$\delta_t = \lambda_A^t - V_A^t \quad (2)$$

$$V_A^{t+1} = V_A^t + \alpha\delta_t \quad (3)$$

Using the recent reinforcement history of a commodity, RL models construct an expected value for a choice alternative (V_A^t) to guide decisions. On each choice, the difference between the expected value and the realized outcome (λ_A) is computed as an RPE. The future value for a commodity (V_A^{t+1}) is then updated according to the RPE (δ) crossed with the learning rate parameter, α, which scales how quickly expected value is updated. Choices for REMI are then predicted to ebb and flow according to relative increases or decreases in the estimated REMI subjective value (see Eq. 6, below). Thus, similar to matching, RL models can help isolate specific effects of REMI through changes in how the magnitude of a reward affects preference (λ) or in the relative weight reward receipt has on subjective value updating for a choice alternative (α).

The application of quantitative choice models to the study of drug choice may also help inform formulation of those models and provide potential insight into decision processes more generally. For instance, non-drug studies of choice are commonly designed as choice between isomorphic commodities (i.e., the same, such as food-food; e.g., Lau and Glimcher 2008), whereas drug choice studies are often allomorphic (i.e., between qualitatively different commodities, like drug/food; e.g., Beckmann et al. 2019). Additionally, drug choice studies commonly employ reward-predictive stimuli (Banks and Negus 2012; Banks and Negus 2017) for drug alternatives that are often absent in isomorphic non-drug choice studies. As such, it remains an open question how to best incorporate allomorphic choice alternatives and the influence of reward-paired stimuli within the context of decision-making models.

One means of quantifying both allomorphic choice and the effects of reward-paired stimuli is through an estimate of substitution for choice alternatives (Beckmann et al. 2019; Green and Freed 1993; Rachlin et al. 1976). Economic substitution commonly refers to the ability of one choice alternative to serve as a (partial) replacement for another alternative as costs vary (e.g., switching to tea if coffee became too expensive). Specific formalizations of both matching (Beckmann et al. 2019) and RL models can quantitatively assess the substitutability of two alternatives by replacing the magnitude of one alternative (such as M_B in Eq. 1) with a free parameter (Ex) that captures the subjective reward magnitude for that alternative relative to other options. For example, under specific REMI-food allomorphic choice conditions, an Ex value of 0.3 would indicate a single food pellet is equivalent to 0.3 μg infusion of REMI. As such, using a free parameter to index the subjective reward magnitude of a commodity can afford one the ability to quantitatively scale conditions that may be qualitatively different, be it through differential reward options or the differential presence of reward-paired cues.

However, use of a parameter to capture subjective reward magnitude in RL models would require reformulation of the traditionally used Eq. 2 from independent subjective value updating to that of a relative ratio (e.g., $\frac{\lambda_A^t}{\lambda_B^t}$), akin to matching in Eq. 1. RL traditionally assumes that subjective value updating for concurrently available alternatives operates independently (i.e., λ_A does not depend on the value of λ_B). Conversely, the use of a ratio term makes the explicit hypothesis that subjective reward value is better expressed as relative to concurrently available alternatives (denoted as the relative value hypothesis). Importantly, these competing hypotheses can be tested through comparing quantitative models with and without the use of the relative ratio updating term to

determine which process is most likely (Wilson and Collins 2019).

Thus, the current study was designed to formulate a dynamic drug choice procedure coupled with quantitative choice modeling to assess the determinants of opioid choice dynamics. Specifically, rats chose between both isomorphic non-drug (food/food) and allomorphic REMI/food commodities within a novel dynamic opioid choice procedure. REMI was chosen because it has comparable reinforcing efficacy to other abused opioids, like fentanyl and heroin (Ko et al. 2002); however, its fast action helps to minimize any direct effects on choice. Across choice blocks, the commodities varied in the presence/absence of reward-predictive cues as well as their relative reward probabilities and magnitudes in an unpredictable manner. Choice data were then analyzed using both matching and RL choice models. It was hypothesized that (1) matching and RL models would be able to successfully formalize the effects of three reward dimensions (probability, magnitude, and associated cues) on both opioid and non-opioid choice; (2) the determinants leading to choice would be similar for both drug and non-drug commodities concordant with evidence that decision-making mechanisms are important for understanding substance abuse; and (3) relative value updating would provide better model fits over independent updating in RL models (i.e., the relative value hypothesis).

Methods

Subjects Twenty-eight male Sprague-Dawley rats (Harlan Inc.; Indianapolis, IN, USA) were used for the experiment as data were collected prior to the NIH sex-as-a-biological-variable mandate. Rats were individually housed on a 12:12-h light:dark cycle (lights on at 7:00 a.m.), had free access to water, and were restricted to approximately 90% of free-feeding body weight. All research was approved by the University of Kentucky Institutional Animal Care and Use Committee (Protocol #2011-0885).

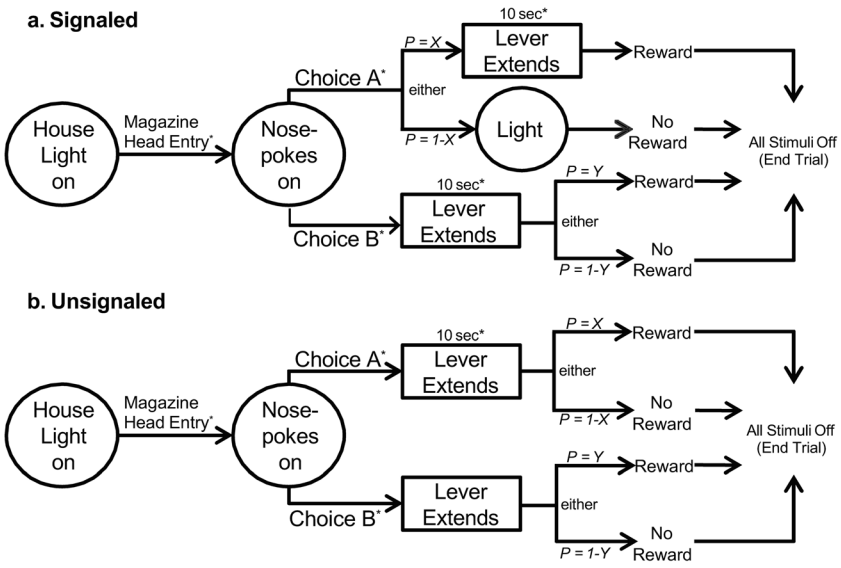
Apparatus Experiments were conducted in Med Associates (St. Albans, VT) conditioning chambers (see supplementary information [SI] for further information).

Initial training Prior to the experiment proper, rats were trained to retrieve food from a central magazine and respond to a nosepoke receptacle for a palatable pellet (see SI). Following initial training, sessions began with the illumination of a house light requiring an orienting response to the central magazine on the front panel. After orienting, either the left or right nosepoke pseudorandomly illuminated on the rear panel. A response to the nosepoke offset the receptacle simultaneously with the presentation of a lever stimulus on the front panel for 10 s. The spatial counterbalancing of the nosepoke-lever combinations was consistent to one side of the chamber from the perspective of the animal. For example, a response to the left nosepoke on the rear panel produced the right lever on the front panel and vice versa for the other alternative. After 10 s, the cue offset (requiring no response) and reward was delivered to the magazine; this delay was constant for all manipulations.

Isomorphic decision-making procedure The probabilistic choice procedure (Lau and Glimcher 2005; Rutledge et al. 2009) was similar to initial training except completion of the orienting response illuminated both nosepokes, allowing for choice between options. Four 30-trial blocks (120 trials total) were pseudorandomly determined with varying reward probabilities. Sessions ended when all trials were completed and full completion was required to be included in the dataset. The possible reward rates, expressed as Option A:B, were 6:1, 2:1, 1:2, and 1:6 (see Fig. 1c for reward probabilities) with both options delivering one food pellet upon payout. The relative reward rate was randomly determined for the first block, and subsequent blocks were selected so that the alternative of greater relative rates switched. 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 condition. On each trial, reward availability was independently calculated for both options and, once programed for delivery, remained until collected.

Rats were also split into the Signaled and Unsignaled groups to test the efficacy of different cue functions. For the Unsignaled group, choice of either option presented the associated lever stimulus for 10 s followed by probabilistic reward delivery according to the current reward probabilities. Importantly, the lever stimuli served only as cues and required no response. For the Signaled group, choice of Option B was identical to the Unsignaled Group. However, choice of Option A produced a lever stimulus *only* when reward followed, while a separate white jewel light above the lever illuminated for an upcoming loss. Thus, the only difference across groups was whether the lever-cue associated with Option A was predictive or probabilistically associated with forthcoming reward (i.e., predictively signaled vs. unsignaled, respectively). All rats completed the initial isomorphic procedure until choice behavior displayed significant sensitivity to the changing relative reward rates (see SI).

Allomorphic decision-making procedure Following the isomorphic procedure, sixteen (eight per group) rats underwent surgery for implantation of a chronic, indwelling catheter to allow for drug self-administration training (see SI). Through training, Option A for both groups became associated with a 3 μg/kg infusion of remifentanil (REMI), an opioid μ receptor agonist (Crespo et al. 2005; Glass et al. 1993), gifted from the National Institute on Drug Abuse (Bethesda, MD, USA).



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

Fig. 1 Sequence of events for the Signaled (**a**) and Unsignaled (**b**) groups. Subsequent manipulations increased the reward magnitude of Option A to two and then three pellets or, in the allomorphic context, from a three to 1 to 10 μg/kg dose of remifentanil in a counterbalanced order. Asterisk "*" 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

Option B continued to deliver one food pellet. The allomorphic experimental procedure was otherwise identical to the isomorphic condition.

Reward magnitude manipulations Rats in the isomorphic and allomorphic choice contexts initially chose between either one food pellet or a 3 μg/kg REMI infusion versus one food pellet at varying probabilities. Subsequently, the magnitude of Option A was increased. For the isomorphic condition, Option A rewarded two and then three food pellets while REMI doses were counterbalanced to either 10 or 1 μg/kg infusions across sessions. Dose was manipulated by changing the pump delivery time (1.77 s/3 μg/kg infusion) and completed simultaneously with the offset of the lever associated with Option A. Each reward magnitude was trained for a maximum of 10 sessions, and training was limited to a 7-session minimum if rats showed significant sensitivity to the changing probabilities prior to the 10-session maximum.

Data analysis Aggregate choice data were calculated as the proportion of choices for Option A as a function of the relative reward rates for Option A over the last 5 days of training for each magnitude condition. Choice data were analyzed using Eq. 1 with the nonlinear mixed effects (NLME) package in R (Pinheiro et al. 2016; Young et al. 2009). However, the reward magnitude for Option B (M_B) was replaced by the scaling constant, *Ex*. *Ex* acts as an exchange rate for the subjective reward magnitude of Option B relative to Option A in units of sucrose pellets or REMI μg/kg infusions for the isomorphic and allomorphic conditions, respectively. For instance, when the reward magnitude of Option A is one pellet, an *Ex* value of 1 suggests the two options were perfectly substitutable. Subject was entered as a nominal random factor, Group as a nominal between-subject fixed factor using dummy coding, and free parameters (S_R, S_M, and *Ex*) were fit to the data. All parameter estimates that are negative in the raw form are shown as absolute values to aid interpretation.

To test the matching assumption of relative valuation at the molecular level, multiple RL models were assessed through model comparison techniques using Akaike Information Criterion (AIC; see SI) on the same set of data. Valuation of Options A and B was first assessed in the Base RL model as shown in Eqs. 2 and 3. Valuation of each commodity was then subsequently made relative in the Scaled Single-Learning model according to Eqs. 4 and 5

$$\delta_A^t = \left(\left(\lambda_A^t/\lambda_B^t\right)^{S_M} - V_A^t\right) \tag{4}$$

$$\delta_B^t = \left(\left(\lambda_B^t/\lambda_A^t\right)^{S_M} - V_B^t\right) \tag{5}$$

where the reward magnitude for both commodities is expressed as a ratio raised to a sensitivity to relative magnitude parameter similar to Eq. 1. The magnitude of Option B (λ_B) was set as the exchange rate obtained from the aggregate

analysis for each group. Value updating then occurred according to Eq. 3

$$V_A^{t+1} = V_A^t + \alpha\delta_A^t$$

where the value for a commodity (V) on the next trial (t) was summated with the RPE from Eqs. 4–5 and scaled according to the learning rate, α. Finally, choices (Ch) for Option A were probabilistically determined according to the softmax Eq. 6.

$$pCh_A^t = \frac{1}{1+\exp\left(-\left[\beta\left(V_A^t - V_B^t\right)\right] + c_A\left(Ch_A^{t-1}\right) + c_B\left(Ch_B^{t-1}\right)\right)} \quad (6)$$

In Eq. 6, the probability of choosing Option A is determined by the relative difference in model-derived value scaled according to the inverse temperature parameter, β, and summated with two perseveration parameters (c_A and c_B). The perseveration parameters weigh the tendency to repeat or alternate from the previous choice independently of model-derived reward value. Values of 1 and 0 were assigned when Option A was previously chosen or disregarded, respectively, and −1 and 0 for Option B. Subsequent models also assessed inclusion of additional α parameters that depended upon the chosen Option [A versus B; α_A, α_B; Scaled Dual-Learning(Option)] or choice outcome [win vs. loss; α_{Win}, α_{Loss}; Scaled Dual-Learning(Outcome)]. RL models were fit to individual rats using the *fmincon* optimization algorithm in MATLAB using maximum likelihood estimation (see SI). Initial parameter values were drawn from 100 sets of uniformly distributed values constrained on the following intervals: $\alpha \in [0, 1]$, $\beta \in [0, 10]$, $c \in [-1, 1]$, and $S_M \in [0, 5]$. Potential parameter differences were assessed using a series of pairwise Wilcoxon signed ranks or rank sign tests with Hochberg error corrections (Hochberg 1988) due to the use of parameter constraints.

Results

Isomorphic choice Figure 2a–b shows the average proportion choices for Option A as a function of the relative reward rates for Option A in the Signaled and Unsignaled group, respectively. Also shown in Fig. 2 are the fits from matching analyses illustrating a successful capturing of choice data. Results from the NLME matching analysis (Fig. 2c) revealed that both the Signaled [S_R = 0.33; S_M = 0.62] and Unsignaled [S_R = 0.29; S_M = 1.06] groups significantly altered their preferences for Option A as the relative reward rates [$F(1, 127) = 213.26$, $p < .001$] and magnitudes [$F(1, 127) = 194.10$, $p < .001$] varied. Stated differently, Option A was more likely to be chosen when reward rates and magnitudes favored Option A and decreased when either dimension became less favorable in an independent manner. Additionally, choices for Option A in the Unsignaled group increased significantly more as its reward magnitude increased [$F(1, 127) = 13.90$, $p < .001$] and Option B for the Unsignaled group had a higher exchange rate [$F(1, 127) = 10.13$, $p = .002$; $Ex = 1.08$] relative to the Signaled group [$Ex = 0.54$]. Expressed in units of sucrose pellets, the exchange rate parameters suggest that an unsignaled food pellet was, as expected, a near perfect substitute for an unsignaled pellet with little bias for either choice option (1 unsignaled pellet was equal to 1.08 unsignaled pellets). Alternatively, an unsignaled food pellet did not substitute as well for a perfectly signaled food pellet (1 perfectly signaled pellet was equal to 0.54 unsignaled pellets). Thus, the subjective reward magnitude for a perfectly signaled pellet was greater than an unsignaled pellet.

RL models successfully parameterized choices (Fig. 2d; Δ\AIC = − 651 from a model predicting chance alone), and model comparisons corroborated the assumptions of the matching equation for both groups. That is, the reformulation of RL value updating to make each commodity value relative (Eqs. 4 and 5) using the matching-derived exchange rates resulted in substantially improved model fits relative to a base (absolute-value updating) RL model [ΔAIC = − 220; see SI]. Further model comparisons revealed adding additional learning rates dependent upon choice outcomes (win versus loss) and option-dependent perseveration parameters (Eq. 6) produced the best AIC values (see SI). Parameter estimates from the best fitting model are shown in Fig. 2e. The Signaled group had significantly greater value updating for loss outcomes than the Unsignaled group [$Z = 2.20$, $p = .028$]. Additionally, inclusion of the matching-derived exchange rates accounted for a group effect a priori, as supported by model AIC improvements.

Allomorphic choice Allomorphic results generally mirrored those from the isomorphic condition. Shown in Fig. 3, the matching NLME model successfully formalized drug choice data and revealed both the Signaled [S_R = 0.29; S_M = 1.12] and Unsignaled [S_R = 0.21; S_M = 1.10] groups were comparably sensitive to changes in relative reward rates [$F(1, 171) = 19.80$, $p < .001$] and magnitudes [$F(1, 171) = 147.38$, $p < .001$] between REMI and food. Although the two groups were not different in sensitivity to changing reward rates or magnitudes, like the isomorphic decision context, the Unsignaled group again had a higher exchange rate [$Ex = 5.86$; $F(1, 171) = 11.90$, $p = .001$] than the Signaled group [$Ex = 3.16$]. The exchange rate parameters suggest that an unsignaled food pellet was a better substitute for an unsignaled REMI infusion (1 pellet was equal to a 5.86 μg/mg unsignaled infusion) than for a signaled REMI infusion (1 pellet was equal to a 3.16 μg/kg signaled infusion). Similar to the isomorphic condition, a signaled REMI infusion had greater subjective reward magnitude as indicated by the unsignaled pellet having a lower Ex

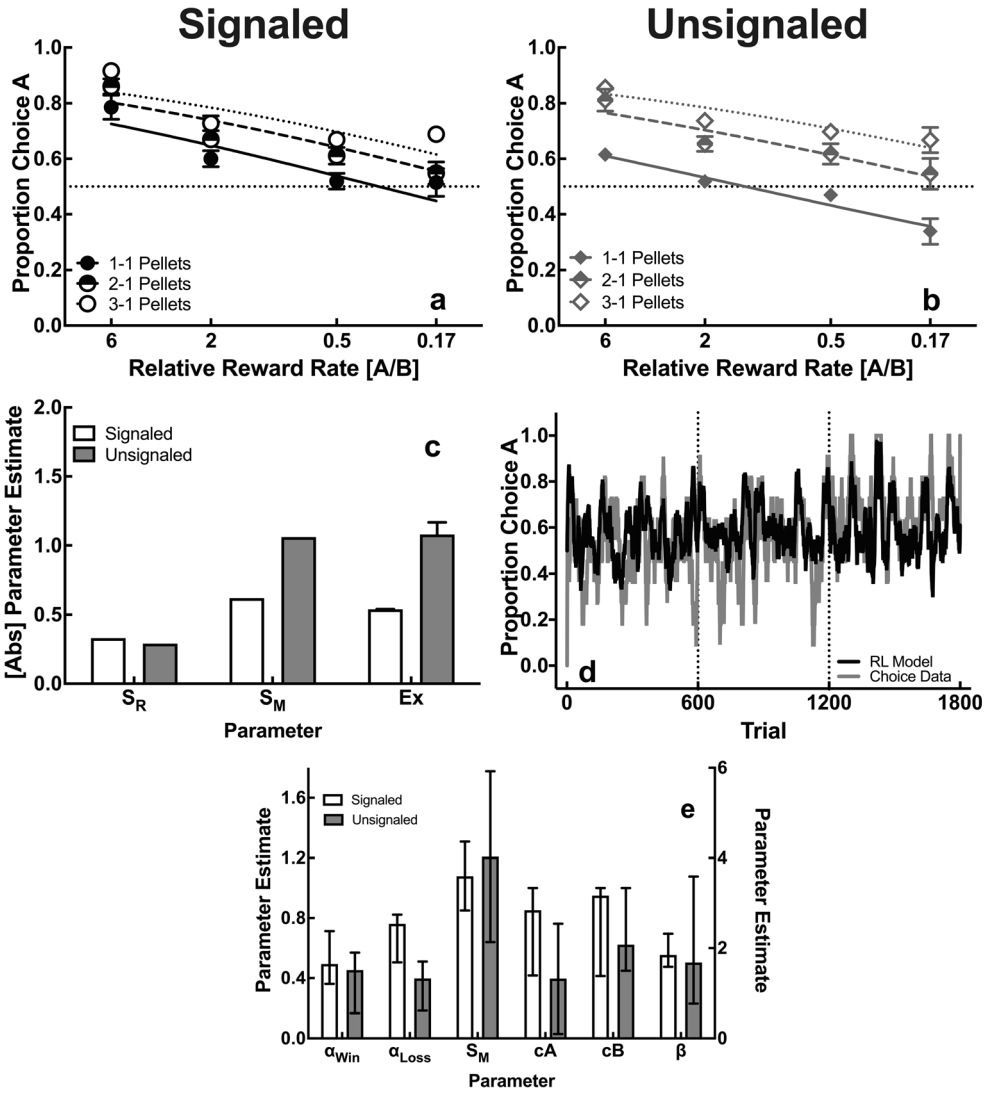


Fig. 2 Mean (± SEM) proportion choice of Option A for the Signaled (**a**) and Unsignaled (**b**) groups as a function of the relative reward rates for Option A in the isomorphic choice context. The x-axis is logged for improved visualization. **c** Mean (± SEM) parameter estimates from the matching model. **d** Example RL model fit. Hatched lines denote changes in the reward magnitude for Option A of 1, 2, and 3 pellets from left to right. **e** Median (± interquartile range) parameter estimates from the RL model. Note: β values are scaled according to the right y-axis. $n = 6$ for panels **a**, **b**, **c**, and **e** for each group

value (i.e., was a worse substitute) relative to the ability of an unsignaled pellet to substitute for an unsignaled REMI infusion.

RL models also successfully parameterized allomorphic drug-associated decision-making (Fig. 3d; ΔAIC = − 640 from a chance model) and corroborated matching model assumptions. Reformulating RL value-updating equations as relative to each commodity in combination with the matching-derived exchange rates (Eqs. 4 and 5) produced substantially improved AIC values relative to a base RL model [ΔAIC = − 348.73; see SI]. Subsequently adding additional learning rate parameters dependent upon choice outcome (win versus loss) and option-dependent perseveration parameters also produced the best overall AIC values. Parameter estimates from the best fitting model are shown in Fig. 3e. RL parameters showed no group effects, but α values for wins were significantly higher than losses across both options and groups [$Z = 3.10$, $p = .002$], and the inclusion of matching-derived exchange values accounted for a group effect a priori, as supported by AIC values.

Discussion

The present study extends the current understanding of drug-associated decision-making in several ways. To our knowledge, the current report is the first to (1) document the effects of reward probability on choices for opioid reward, (2) use a dynamic, probabilistic choice procedure to assess both drug and non-drug decision-making, (3) demonstrate that cues signaling impending reinforcement (i.e., prior to reward delivery) can modulate drug as well as non-drug choices (cf. Chow et al. 2017; Smith et al. 2018; Zentall 2016), (4) successfully apply RL and matching models to opioid-associated decision making, and (5) show that choice-by-choice decisions were better described by relative value updating in RL models (consistent with matching) than the traditional independent value updating assumption. Overall, behavioral data showed a striking similarity in decision-making mechanisms between drug and non-drug choice,

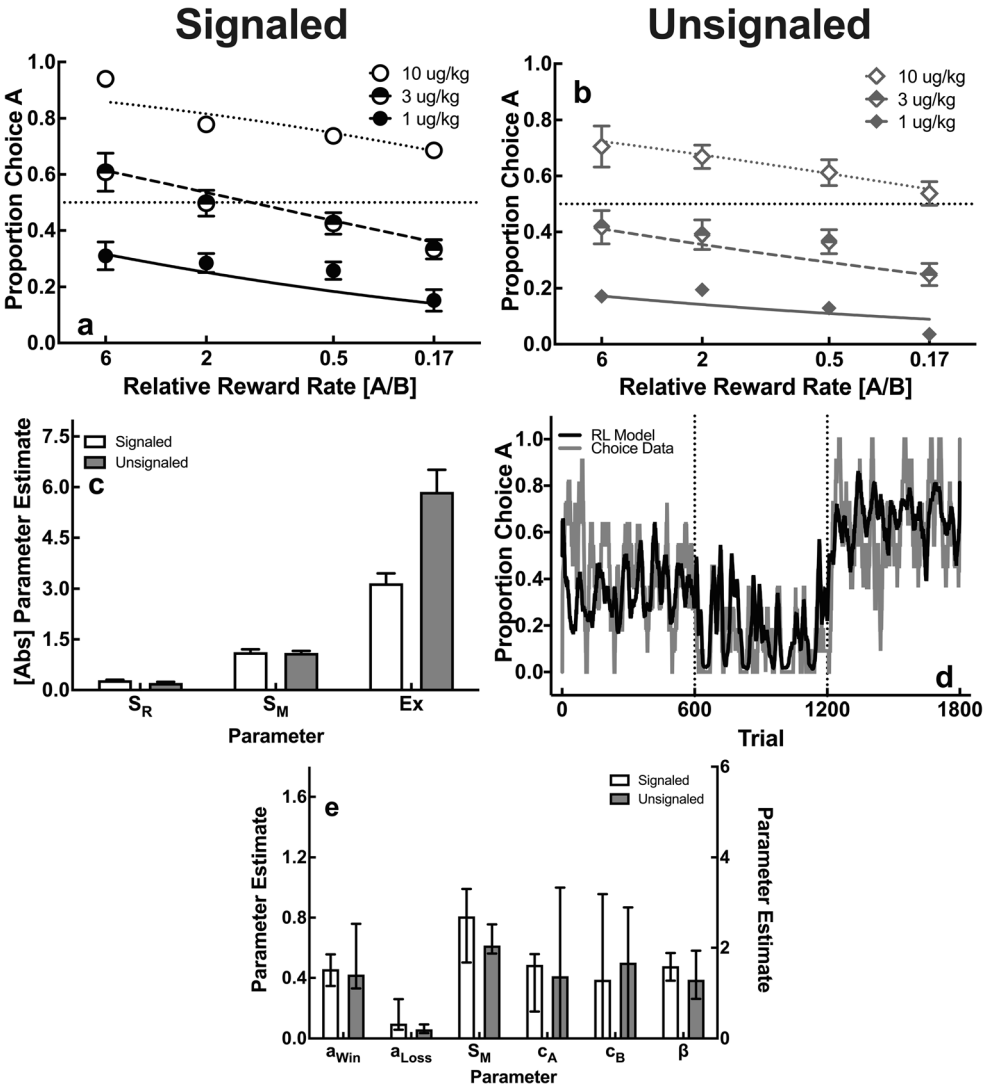


Fig. 3 Mean (± SEM) proportion choice of Option A for the Signaled (**a**) and Unsignaled (**b**) groups as a function of the relative reward rates for Option A in the allomorphic choice context. The x-axis is logged for improved visualization. **c** Mean (± SEM) parameter estimates from the matching model. **d** Example RL model fit. Hatched lines denote changes in the reward magnitude for Option A of 3, 1, and 10 μg/kg from left to right. **e** Median (± interquartile range) parameter estimates from the RL model. Note: β values are scaled according to the right y-axis. $n = 8$ for panels **a**, **b**, **c**, and **e** for each group

suggesting that SUDs may stem from modulations of relative value (Heyman 2013; Hogarth 2020; Hogarth and Field 2020).

Across both choice contexts (isomorphic and allomorphic), Option A was the generally preferred commodity when the relative reward rates and magnitudes favored it and decreased as magnitudes and probabilities favored Option B. Thus, the present work corroborates previous drug choice studies showing drug-associated decision making is sensitive to various reward-relevant dimensions known from work with non-drug choice (e.g., Banks and Negus 2012; Beckmann et al. 2019; Moeller and Stoops 2015; Woolverton and Rowlett 1998). Although the preference shifts across magnitudes tended to be larger in the allomorphic condition, this is likely attributable to the scale on which REMI doses were manipulated relative to food pellets (linear vs. logarithmic). Within the isomorphic condition, increasing food reward by a greater amount may have produced similar preference switches as the allomorphic condition. Additionally, shifting position on one relative reinforcer dimension (e.g., magnitude) to a sufficient degree had the potential to exceed the effects of the other dimension (probability). Specifically, increasing the food reward magnitude of Option A to 3 pellets influenced preferences to such a degree that even a 1:6 reward ratio favoring Option B could not induce a preference reversal. Conversely, decreasing the REMI dose of Option A to 1 μg/kg changed the relative choice context such that even a 6:1 reward rate favoring REMI did not produce REMI preference. The above effects therefore warrant caution in making statements about absolute commodity preferences that could stem from insufficiently altering reward-relevant dimensions (e.g., not increasing relative magnitude or probability sufficiently, especially from manipulations that include only a single dose and/or probability).

Quantitative choice models were also able to capture the changes in choice across both contexts. That matching (Eq. 1) could describe drug-associated decision-making corroborates previous research using cocaine (Anderson et al. 2002;

Beckmann et al. 2019; Hutsell et al. 2015), but the current study is the first to apply matching and RL to choices for opioid reward. Additionally, the current results suggest that value updating in RL models might be better expressed as relative to other concurrently available commodities. In both choice contexts, reformulating the RL reward magnitude term as a scaled, relative term akin to matching models (Davison and McCarthy 1988) resulted in substantially improved model fits. The confirmation of the relative value hypothesis is important, as having the correct model architecture should improve the ability to detect potential neurological correlates of reward value (Corrado et al. 2009). For instance, many studies are interested in determining whether all reward value reduces to a "common currency" neuronal signal within the brain (Levy and Glimcher 2012). The current findings suggest value is context-dependent, and thus any measurement of reward value would not be absolute. That said, further research assessing relative versus absolute RL value updating in allomorphic decisions is needed, as this is, to our knowledge, currently the only report to do so.

If the value of a commodity is indeed relative to other available options, then studies examining commodity value should necessarily provide manipulatable alternatives to better understand the value metric. For instance, some models of reward value offer only one manipulated commodity as a standard protocol (e.g., single-schedule substance abuse models; Banks and Negus 2012) or assume that reward value is modulated by a single dimension (e.g., expected value and unit price; Hursh and Silberberg 2008). However, the present results corroborate previous work (Beckmann et al. 2019; Davison and McCarthy 1988; Hutsell et al. 2015; Moeller and Stoops 2015) in demonstrating that reward value changes depend on other concurrently available options and can be independently modulated by several reward dimensions (e.g., reward magnitude and probability). Other reward dimensions not manipulated here such as effort, delay to reward, and even state-dependent effects like negative affect can also influence drug choice (Canchy et al. 2020; Eldar and Niv 2015; Hogarth et al. 2015; Mitchell 2004) but can be similarly incorporated within the theoretical framework of independently operating reward dimensions combining to affect choice (Davison and McCarthy 1988; Rachlin 1971). As further evidence, additional quantitative models that assume unidimensional reward spaces (e.g., expected value) were assessed using the current dataset; these models resulted in poorer fits of the current data relative to the assumptions of matching (see SI). Thus, models of reward value, including drug value, should at the very least include alternative commodities in their experiments to properly test valuation.

The current results also demonstrate that the inclusion of stimuli differentially predictive of reward can alter the scaling of relative reward magnitude effects on choice. Across both choice contexts, an unsignaled food pellet or REMI reward was a poor substitute for a perfectly signaled (i.e., 100% of the time) reward that led to subsequent increased preference for Option A (i.e., the Signaled option). Such a finding is important as it again points to a similarity in choice mechanisms between drug and non-drug choice (cf. McDevitt et al. 2016; Zentall 2016), and it suggests that contexts with certain arrangements of reward-associated cues can greatly increase the value of drug alternatives relative to non-drug alternatives.

RL analyses also showed that the Signaled group weighed loss outcomes as larger than the Unsignaled group within the isomorphic choice context, and both groups weighed wins more heavily than losses in the allomorphic choice context. The former result is consistent with the reduced sensitivity to magnitude for the Signaled group at the aggregate level of analysis. That is, the increased salience of reward occurrence (or losses) may have overshadowed reward magnitude changes and led to increased weighting of losses relative to the Unsignaled group on a choice-by-choice basis. The reason for wins having greater weight than losses during opioid allomorphic decision-making is currently unclear. One possibility may be that REMI increases or decreases the relative salience of wins or losses, respectively. However, wins and losses having a differential impact on decision-making does corroborate previous research (Glimcher et al. 2013; Marshall and Kirkpatrick 2017; Rutledge et al. 2009), and the results highlight the ability of quantitative models to isolate specific facets of how the decision-making process may change across contexts.

The current results also potentially have broader implications for the treatment of SUD as a disorder of relative reward valuation. Across tested doses, rats showed significant reductions in REMI choices as a function of reducing reward probabilities and magnitudes. Matching and RL models were also successfully able to parameterize drug-associated decision-making similar to isomorphic food choices (see Tables S3-4). Therefore, the collective results suggest that decision making, even for drugs, was value-based across all contexts. Importantly, the implications of drug-associated decision making being ostensibly value based validates treatments that alter the relative value of drugs. For instance, contingency management treatments offer reward for drug abstinence (Davis et al. 2016), and potential psychometric measures of non-drug-associated reward have been developed (Acuff et al. 2019). Additionally, an understanding of how structural factors (i.e., the context of an individual) contribute to individual drug use will likely facilitate balancing non-drug-derived reinforcement (Lee et al. 2018). The present results also inform conceptual models of SUDs. For instance, the value-based models herein contrast somewhat with various SUD models suggesting drug use is compulsive or habitual (Everitt and Robbins 2016; Kalivas et al. 2005; Redish 2004; Vandaele and Janak 2018). Evidence for compulsive, habitual drug responding (even when choice-specific

perseveration parameters were included in model analyses) was not found in the current study. Rather, behavioral shifts in line with the predictions of value-based quantitative choice models were found. As such, the current study corroborates previous notions of conceptualizing SUDs as an extension of similar decision-making systems for non-drug rewards (Banks and Negus 2012; Heyman 2013; Hogarth 2018; Hogarth 2020; Hogarth and Field 2020; Negus and Banks 2018; Rachlin 2007) and also illustrates that making generalized statements about the relative abuse liability of drugs is difficult due to the highly context-dependent nature of drug value.

Nonetheless, one important variable left to future research is the impact of drug use history. Previous research has shown that increased drug use history may increase preferences for opioid rewards, particularly during withdrawal (Lenoir et al. 2013; Negus and Banks 2018; Wade-Galuska et al. 2011). Opioid use history remains an open research question that can be tested thoroughly utilizing the models provided herein. For instance, assessment of how RL or matching parameters vary during different stages of drug use, such as withdrawal, can help dissociate between value-based and habitual-responding hypotheses. Large reductions in matching sensitivity values or reduced alpha values in combination with increased perseveration parameters in RL models may provide evidence of decreased value-based decision-making consistent with habit-like responding. Alternatively, it is also possible that during withdrawal value-based processes are modulated (Hogarth 2020). For example, opioid withdrawal may modulate exchange rates to suitably favor drug reward, increase reward dimension sensitivities, or elevate drug-option alpha values in RL models to produce what could appear as compulsive responding (e.g., the 10 μg/kg REMI condition herein), yet substantiate the retention of value-based decision making. A further test of interest would be whether changes in any parameter values persist after prolonged abstinence.

Potential limitations of the study are worth noting. First, the current study only used male rats, which will require future research to identify if the findings generalize to female rats. Second, REMI, an opioid μ receptor agonist with brief reinforcing kinetics, was used that may have a lower abuse liability compared to other μ agonists (Baylon et al. 2000). However, the reinforcing efficacy of REMI has been documented as comparable to other μ receptor agonists such as fentanyl and heroin (Ko et al. 2002), and there are documented cases of REMI abuse (Levine and Bryson 2010). The current study also did not manipulate drug use history to assess if longer histories may produce more compulsive-like behavior. Finally, all rats first underwent isomorphic food-food training prior to REMI-food choice which may have influenced the results; this procedure was employed to ensure the dynamic choice task was learned prior to introducing REMI. If future research chooses to exclude the initial food-food training, care should be taken that any differences seen in drug conditions are not simply due to an insufficient learning of the dynamic procedure. Overall, dynamic drug choice procedures like those used herein combined with quantitative choice modeling provide a powerful, unique avenue for testing competing hypotheses regarding underlying processes that govern substance use disorder.

In conclusion, the present study highlights, at multiple levels of analysis, that the quantitative assumptions of choice as relative, multidimensional, and context dependent are substantiated in both food- and opioid-associated decision-making. At the aggregate level, preference for a commodity was shown to be determined by its reward rate, magnitude, and the presence of reward-predictive cues relative to another commodity. Additionally, the common assumption of independent value updating in RL models was improved when made relative to concurrent alternatives. Future research may expand the present findings through assessing allomorphic choice of different drugs of abuse, the influence of different pharmacological treatment strategies (Banks et al. 2015), potential risk factors (e.g., prolonged drug use history), and identification of underlying neural pathways. Broader implications from the current research also point to the need for substance abuse models to include more complex, dynamic, and multidimensional procedures to better identify determinants of drug value and its underlying mechanisms. Collectively, the present results suggest that economic variables (e.g., substitute availability; Davis et al. 2016) should be considered potential treatment options for SUDs, as well as important variables to consider during the development of novel SUD therapies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00213-021-05830-x.

Acknowledgements The authors would like to thank Josh Lavy for his technical assistance and Andrew T. Marshall for his input on RL model analysis.

Author contribution APS and JSB made substantial contributions to the conception, design, collection, and analysis of the work as well as drafting, revising, and approving the final version. JSB agrees to be the point of contact for questions related to the current work.

Funding The present research was supported by funding from the National Institute on Drug Abuse R01DA047368, T32DA016176, and T32DA07304.

References

Acuff SF, Dennhardt AA, Correia CJ, Murphy JG (2019) Measurement of substance-free reinforcement in addiction: a systematic review. Clin Psychol Rev 70:79–90

Ahmed SH (2005) Imbalance between drug and non-drug reward availability: a major risk factor for addiction. Eur J Pharmacol 526:9–20

Ahmed SH (2010) Validation crisis in animal models of drug addiction: beyond non-disordered drug use toward drug addiction. Neurosci

Biobehav Rev 35:172–184. https://doi.org/10.1016/j.neubiorev.2010.04.005
Anderson KG, Velkey AJ, Woolverton WL (2002) The generalized matching law as a predictor of choice between cocaine and food in rhesus monkeys. Psychopharmacology 163:319–326
Banks ML, Hutsell BA, Schwienteck KL, Negus SS (2015) Use of preclinical drug vs. food choice procedures to evaluate candidate medications for cocaine addiction. Curr Treat Opt Psych 2:136–150
Banks ML, Negus SS (2012) Preclinical determinants of drug choice under concurrent schedules of drug self-administration. Adv Pharmacol Sci 2012:281768. https://doi.org/10.1155/2012/281768
Banks ML, Negus SS (2017) Insights from preclinical choice models on treating drug addiction. Trends Pharmacol Sci 38:181–194
Baylon GJ, Kaplan HL, Somer G, Busto UE, Sellers EM (2000) Comparative abuse liability of intravenously administered remifentanil and fentanyl. J Clin Psychopharmacol 20:597–606
Beckmann JS, Chow JJ, Hutsell BA (2019) Cocaine-associated decision-making: toward isolating preference. Neuropharmacology 153:142–152
Canchy L, Girardeau P, Durand A, Vouillac-Mendoza C, Ahmed SH (2020) Pharmacokinetics trumps pharmacodynamics during cocaine choice: a reconciliation with the dopamine hypothesis of addiction. Neuropsychopharmacol 46:288–296
Chow JJ, Smith AP, Wilson AG, Zentall TR, Beckmann JS (2017) Suboptimal choice in rats: incentive salience attribution promotes maladaptive decision-making. Behav Brain Res 320:244–254. https://doi.org/10.1016/j.bbr.2016.12.013
Corrado GS, Sugrue LP, Brown JR, Newsome WT (2009) The trouble with choice: studying decision variables in the brain. In: Neuroeconomics. Elsevier, Amsterdam, pp 463–480
Crespo JA, Sturm K, Saria A, Zernig G (2005) Simultaneous intra-accumbens remifentanil and dopamine kinetics suggest that neither determines within-session operant responding. Psychopharmacology 183:201–209. https://doi.org/10.1007/s00213-005-0180-7
Davis DR, Kurti AN, Skelly JM, Redner R, White TJ, Higgins ST (2016) A review of the literature on contingency management in the treatment of substance use disorders, 2009-2014. Prev Med 92:36–46. https://doi.org/10.1016/j.ypmed.2016.08.008
Davison M, McCarthy D (1988) The matching law: a research review. Lawrence Erlbaum Associates Inc., New Jersey
Eldar E, Niv Y (2015) Interaction between emotional state and learning underlies mood instability. Nat Commun 6(1):1–10
Everitt BJ, Robbins TW (2016) Drug addiction: updating actions to habits to compulsions ten years on. Annu Rev Psychol 67:23–50
Glass P, Hardman D, Kamiyama Y, Quill TJ, Marton G, Donn KH, Grosse CM, Hermann D (1993) Preliminary pharmacokinetics and pharmacodynamics of an ultra-short-acting opioid: remifentanil (GI87084B). Anesth Analg 77:1031–1040
Glimcher PW (2011) Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc Natl Acad Sci U S A 108:15647–15654
Glimcher PW, Camerer CF, Fehr E, Poldrack RA (2013) Neuroeconomics: decision making and the brain. Academic Press, Cambridge
Green L, Freed DE (1993) The substitutability of reinforcers. J Exp Anal Behav 60:141–158
Heather N (2017) Is the concept of compulsion useful in the explanation or description of addictive behaviour and experience? Addict Behav Rep 6:15–38
Heyman GM (2013) Addiction: an emergent consequence of elementary choice principles. Inquiry 56:428–445
Hochberg Y (1988) A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75:800–802
Hogarth L (2018) A critical review of habit theory of drug dependence. In: The Psychology of Habit. Springer, Berlin, pp 325–341
Hogarth L (2020) Addiction is driven by excessive goal-directed drug choice under negative affect: translational critique of habit and compulsion theory. Neuropsychopharmacology 45:720–735
Hogarth L, Field M (2020) Relative expected value of drugs versus competing rewards underpins vulnerability to and recovery from addiction. Behav Brain Res 394:112815
Hogarth L et al (2015) Negative mood reverses devaluation of goal-directed drug-seeking favouring an incentive learning account of drug dependence. Psychopharmacology 232:3235–3247
Hursh SR, Silberberg A (2008) Economic demand and essential value. Psychol Rev 115:186
Hutsell BA, Negus SS, Banks ML (2015) A generalized matching law analysis of cocaine vs. food choice in rhesus monkeys: effects of candidate 'agonist-based' medications on sensitivity to reinforcement. Drug Alcohol Depend 146:52–60
Kalivas P, Volkow N, Seamans J (2005) Unmanageable motivation in addiction: a pathology in prefrontal-accumbens glutamate transmission. Neuron 45:647–650
Ko M, Terner J, Hursh S, Woods J, Winger G (2002) Relative reinforcing effects of three opioids with different durations of action. J Pharmacol Exp Ther 301:698–704
Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 84:555–579
Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463
Lee JO, Cho J, Yoon Y, Bello MS, Khoddam R, Leventhal AM (2018) Developmental pathways from parental socioeconomic status to adolescent substance use: alternative and complementary reinforcement. J Youth Adoles 47:334–348
Lenoir M, Cantin L, Vanhille N, Serre F, Ahmed SH (2013) Extended heroin access increases heroin choices over a potent nondrug alternative. Neuropsychopharmacology 38:1209
Levine AI, Bryson EO (2010) Intranasal self-administration of remifentanil as the foray into opioid abuse by an anesthesia resident. Anesth Analg 110:524–525
Levy DJ, Glimcher PW (2012) The root of all value: a neural common currency for choice. Curr Opin Neurobiol 22:1027–1038
Marshall AT, Kirkpatrick K (2017) Reinforcement learning models of risky choice and the promotion of risk-taking by losses disguised as wins in rats. J Exper Psychol Animal Learn Cogn 43:262
McDevitt MA, Dunn RM, Spetch ML, Ludvig EA (2016) When good news leads to bad choices. J Exp Anal Behav 105:23–40. https://doi.org/10.1002/jeab.192
Mitchell SH (2004) Effects of short-term nicotine deprivation on decision-making: delay, uncertainty and effort discounting. Nicotine Tob Res 6:819–828
Moeller SJ, Stoops WW (2015) Cocaine choice procedures in animals, humans, and treatment-seekers: can we bridge the divide? Pharmacol Biochem Behav 138:133–141
Negus SS, Banks ML (2018) Modulation of drug choice by extended drug access and withdrawal in rhesus monkeys: implications for negative reinforcement as a driver of addiction and target for medications development. Pharmacol Biochem Behav 164:32–39. https://doi.org/10.1016/j.pbb.2017.04.006
Pinheiro J, Bates D, DebRoy S, Team RC (2016) *nlme: linear and nonlinear mixed effects models* R package version 31-128
Rachlin H (1971) On the tautology of the matching law. J Exp Anal Behav 15:249–251
Rachlin H (2007) In what sense are addicts irrational? Drug Alcohol Depend 90:S92–S99
Rachlin H, Green L, Kagel JH, Battalio RC (1976) Economic demand theory and psychological studies of choice. In: Psychology of Learning and Motivation, vol 10. Elsevier, Amsterdam, pp 129–154
Redish AD (2004) Addiction as a computational process gone awry. Science 306:1944–1947

Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW (2009) Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. J Neurosci 29:15104–15114

Smith AP, Hofford RS, Zentall TR, Beckmann JS (2018) The role of 'jackpot' stimuli in maladaptive decision-making: dissociable effects of D1/D2 receptor agonists and antagonists. Psychopharmacology 235:1–11

Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304:1782–1787

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction, vol 1. MIT press, Cambridge

Vandaele Y, Janak PH (2018) Defining the place of habit in substance use disorders. Prog Neuro-Psychopharmacol Biol Psychiatry 87:22–32

Wade-Galuska T, Galuska CM, Winger G (2011) Effects of daily morphine administration and deprivation on choice and demand for remifentanil and cocaine in rhesus monkeys. J Exp Anal Behav 95:75–89

Wilson RC, Collins AG (2019) Ten simple rules for the computational modeling of behavioral data. Elife 8:e49547

Woolverton WL, Rowlett JK (1998) Choice maintained by cocaine or food in monkeys: effects of varying probability of reinforcement. Psychopharmacology 138:102–106

Young ME, Clark M, Goffus A, Hoane MR (2009) Mixed effects modeling of Morris water maze data: advantages and cautionary notes. Learn Motiv 40:160–177

Zentall TR (2016) Resolving the paradox of suboptimal choice. J Exper Psychol Animal Learn Cogn 42:1–14