Audiovisual cues must be predictable and win-paired to drive risky choice

Brett A. Hathaway[1], Dexter R. Kim[1]*, Salwa B. A. Malhas[1]*, Kelly M. Hrelja[1], Lauren Kerker[1], Tristan J. Hynes[1], Cailean B. W. Harris[1], Angela J. Langdon[2], & Catharine A. Winstanley[1]

[1]*Department of Psychology, Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, BC, Canada*

[2]*Neural Computations in Learning, National Institute of Mental Health, Bethesda, MD, USA*

*These authors contributed equally to this work

Correspondence should be addressed to Brett A. Hathaway or Dr. Catharine A. Winstanley

Address: Djavad Mowafaghian Centre for Brain Health,
Department of Psychology,
2215 Wesbrook Mall
Vancouver, BC, V6T 1Z3,
Canada

Email: bretthathaway@psych.ubc.ca or cwinstanley@psych.ubc.ca

Tel: 604-827-5083 or 604-822-3128

Acknowledgements: This work was supported by a Discovery Grant awarded to CAW from the Natural Sciences and Engineering Council of Canada (NSERC; RGPIN-2017-05006) and a project grant (PJT-162312) awarded to CAW from the Canadian Institutes for Health Research (CIHR). BAH was supported by a CIHR Doctoral Award, and DRK was supported by an Undergraduate Summer Research Award from NSERC. The experimental work took place at a UBC campus situated on the traditional, ancestral, and unceded land of the xʷməθkʷəy̓əm (Musqueam), sə̓lílwətaʔɬ/Selilwitulh (Tsleil-Waututh) and Sḵwx̱wú7mesh (Squamish) Peoples. We acknowledge and are grateful for their stewardship of this land for thousands of years.

Abstract

Risky or maladaptive decision making is thought to be central to the etiology of both drug and gambling addiction. Win-concurrent cues can enhance disadvantageous, risky choice in both rats and humans, yet it is unclear which aspects of the cue-reward contingencies drive this effect. Here, we implemented six variants of the rat Gambling Task (rGT), in which animals can maximise their sugar pellet profits by avoiding options paired with higher per-trial gains but disproportionately longer and more frequent time-out penalties. When audiovisual cues were delivered concurrently with wins, and scaled in salience with reward size, significantly more rats preferred the risky options as compared to the uncued rGT. Similar results were observed when the relationship between reward size and cue complexity was inverted, and when cues were delivered concurrently with all outcomes. Conversely, risky choice did not increase when cues occurred randomly on 50% of trials, and decision making actually improved when cues were coincident with losses alone. As such, cues do not increase risky choice by simply elevating arousal, or amplifying the difference between wins and losses. It is instead important that the cues are reliably associated with wins; presenting the cues on losing outcomes as well as wins does not diminish their ability to drive risky choice. Computational analyses indicate that win-paired cues reduced the impact of losses on decision making. These results may help us understand how sensory stimulation can increase the addictive nature of gambling and gaming products.

The lights and sounds of a casino are physiologically arousing and increase enjoyment, particularly in those with pathological gambling (Dixon et al., 2014; Loba et al., 2001; Spetch et al., 2020). Indeed, gambling-related cues can cause intense cravings in such individuals, and there is increasing concern over their contribution to addiction (Limbrick-Oldfield et al., 2017; Alter, 2018). Such cues feature prominently in electronic gaming machines (EGMs), which are specifically designed to encourage excessive gambling (Griffiths, 1993). In particular, the salient lights and sounds associated with EGMs are thought to facilitate problematic gambling in susceptible individuals and lead to an increased state of immersion as well as overestimation of the number of wins (Dixon et al., 2014; Alter, 2018; Murch et al., 2017). Deficits in cost/benefit decision making are particularly pronounced in individuals who prefer EGMs over other forms of gambling (Goudriaan et al., 2005). Together, this evidence suggests that cue-induced impairments in cost/benefit decision making may be a critical risk factor for the development and maintenance of behavioural addictions such as gambling disorder. However, the impact of salient audiovisual cues on decision making has not been well characterized.

One approach to investigate the influence of cues in cost/benefit decision making utilizes the rat Gambling Task (rGT), a rodent analog of the human Iowa Gambling Task (IGT, Zeeb et al., 2009; Bechara et al., 1994). In both tasks, optimal performance is attained by avoiding the two high-risk, high-reward options and instead favouring the low-risk options associated with lower per-trial gains. On the rGT, these low-risk, low-reward options result in less frequent and shorter time-out penalties and therefore more sucrose pellets may be earned overall. The addition of reward-concurrent audiovisual cues leads to a higher proportion of rats establishing a disadvantageous risky decision-making profile (Barrus & Winstanley, 2016). A similar effect of reward-paired cues on risky decision making has been observed in humans (Cherkasova et al., 2018). Such cues also appear to lead to inflexibility in decision-making patterns, as indicated by insensitivity to reinforcer devaluation in the cued but not the uncued rGT (Hathaway et al., 2021; Zeeb & Winstanley, 2013).

Investigating the learning dynamics of the uncued versus cued rGT using a series of reinforcement learning models revealed that potentiated learning from the cued rewards does not drive risk preference on the cued rGT, as might be expected (Langdon et al., 2019). Instead, rats on the cued task were relatively insensitive to the time-out penalties, particularly for the risky options featuring lengthy and more frequent penalties. This was indicated by differences in parameters governing learning about losses between the uncued and cued tasks across all models tested.

Several theories could explain these results. One possibility is that higher levels of arousal resulting from exposure to the cues persists through the time-out penalties on subsequent trials and thereby alters the processing of the punishment signal. Alternatively, reward-paired cues may change the representation of the task structure in prefrontal cortices, such that punishments are not correctly integrated into the stored action-outcome contingencies as the rats learn to choose between the options. It is possible that the salient cues cause rats to represent winning outcomes as different “states” than losing outcomes, and learning about one state does not generalize to another (Niv, 2019). In that case, time-out penalties would not appropriately devalue the risky options and rats would tend to choose the options offering the highest per-trial reward.

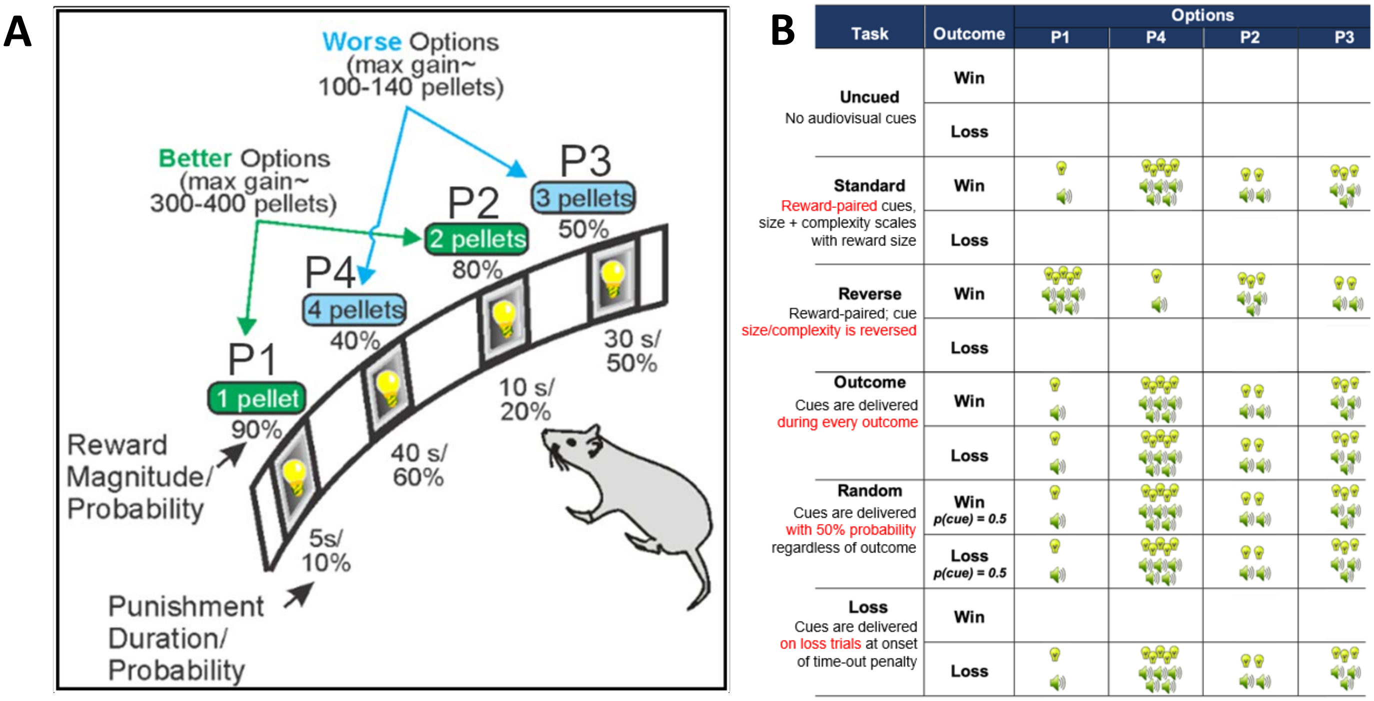
To test these theories and further investigate impairments in risky decision making induced by reward-paired cues, we designed several variants of the rGT that varied the size and position of the cues in the task. In the standard-cued task, the audiovisual cues scale in magnitude and complexity with reward size. To determine whether this scaling is a necessary feature to drive risky choice, we implemented an inverse relationship between cue complexity/magnitude and reward size. We next tested whether cuing all outcomes, ostensibly making trial outcomes more similar and perhaps permitting correct integration into each option’s stored value, would similarly impact risky choice as solely cuing the wins. To test whether increased sensory stimulation is sufficient to increase risky choice, cues were played randomly on 50% of trials, regardless of outcome. Lastly, we paired cues with losses instead of wins to investigate whether win-paired cues are necessary to drive risky choice. A reinforcer devaluation procedure was also utilized at the end of training to determine which cue-outcome associations would lead to inflexibility in choice. Reinforcement learning models were used to identify differences in the learning dynamics early in training that may underlie differential choice patterns across the tasks.

Results

Baseline behaviour

Choice

Figure 1A depicts a schematic of the rGT, and the cue variants are described in 2B. Differences in decision making induced by the different cue variants were assessed by comparing the percent choice of the four options (optimal: P1, P2; risky: P3, P4) at the end of training, once a statistically stable baseline was reached. When comparing P1-P4 choice in an omnibus ANOVA, a significant choice x task interaction was observed ($F(14,415) = 2.16$, $p = .009$; see Figure 2A). Group comparisons for all rats can be found in Table 1. Generally speaking, differences in P1-P4 choice were found between task versions featuring win-paired cues (standard-cued, reverse-cued, outcome-cued) and those without (uncued, random-cued, loss-cued). These differences may have been driven by risk-preferring rats, as a significant choice x task x risk status interaction was also observed (F(11,415) = 2.76, p = .002), and only risk-preferring rats exhibited task differences (risk-preferring: F(12,96) = 1.83, p = .05; optimal: F(10,248) = 1.12, p = .35). However, only one *post-hoc* comparison reached marginal significance among risky rats, likely due to the relatively low number of risk-preferring rats for some task variants.



Task	Outcome	Options P1	P4	P2	P3
Uncued No audiovisual cues	**Win**				
	Loss				
Standard Reward-paired cues, size + complexity scales with reward size	**Win**	💡 🔊	💡💡💡💡💡 🔊🔊🔊🔊🔊	💡💡 🔊🔊	💡💡💡 🔊🔊🔊
	Loss				
Reverse Reward-paired; cue size/complexity is reversed	**Win**	💡💡💡💡💡 🔊🔊🔊🔊🔊	💡 🔊	💡💡💡 🔊🔊🔊	💡💡 🔊🔊
	Loss				
Outcome Cues are delivered during every outcome	**Win**	💡 🔊	💡💡💡💡💡 🔊🔊🔊🔊🔊	💡💡 🔊🔊	💡💡💡 🔊🔊🔊
	Loss	💡 🔊	💡💡💡💡💡 🔊🔊🔊🔊🔊	💡💡 🔊🔊	💡💡💡 🔊🔊🔊
Random Cues are delivered with 50% probability regardless of outcome	**Win** *p(cue) = 0.5*	💡 🔊	💡💡💡💡💡 🔊🔊🔊🔊🔊	💡💡 🔊🔊	💡💡💡 🔊🔊🔊
	Loss *p(cue) = 0.5*	💡 🔊	💡💡💡💡💡 🔊🔊🔊🔊🔊	💡💡 🔊🔊	💡💡💡 🔊🔊🔊
Loss Cues are delivered on loss trials at onset of time-out penalty	**Win**				
	Loss	💡 🔊	💡💡💡💡💡 🔊🔊🔊🔊🔊	💡💡 🔊🔊	💡💡💡 🔊🔊🔊

Fig. 1A: Task schematic of the cued rGT. A nose poke response in the food tray extinguished the traylight and initiated a new trial. After an inter-trial interval (ITI) of 5 s, four stimulus lights were turned on in holes 1, 2, 4, and 5, each of which was associated with a different number of sugar pellets. The order of the options from left to right was counter-balanced within each cohort to avoid development of a simple side bias (version A (shown): P1, P4, P3, P2; version B: P4, P1, P3, P2). The animal was required to respond at a hole within 10 s. This response was then rewarded or punished depending on the reinforcement schedule for that option. If the animal lost, the stimulus light in the chosen hole flashed at a frequency of 0.5 Hz for the duration of the time-out penalty, and all other lights were extinguished. The maximum number of pellets available per 30 min session shows that P1 and P2 are more optimal than P3 and P4. The percent choice of the different options is one of the primary dependent variables. A score variable is also calculated, as for the IGT, to determine the overall level of risky choice as follows: [(P1 + P2) – (P3 + P4)]. Figure is modified from Barrus and Winstanley (2016). Fig. 2B describes the 6 variants of the rGT. On the uncued variant, no audiovisual cues were present. The standard task featured audiovisual cues that scaled in complexity and magnitude with reward size. The reverse-cued variant inverted this relationship, such that the simplest cue was paired with the largest reward, and vice versa. Audiovisual cues were paired with both wins and losses for the outcome-cued variant. For the random-cued variant, cues were played on 50% of trials, regardless of outcome. Lastly, for the loss-cued variant, cues were only paired with losing outcomes, at the onset of the time-out penalty.

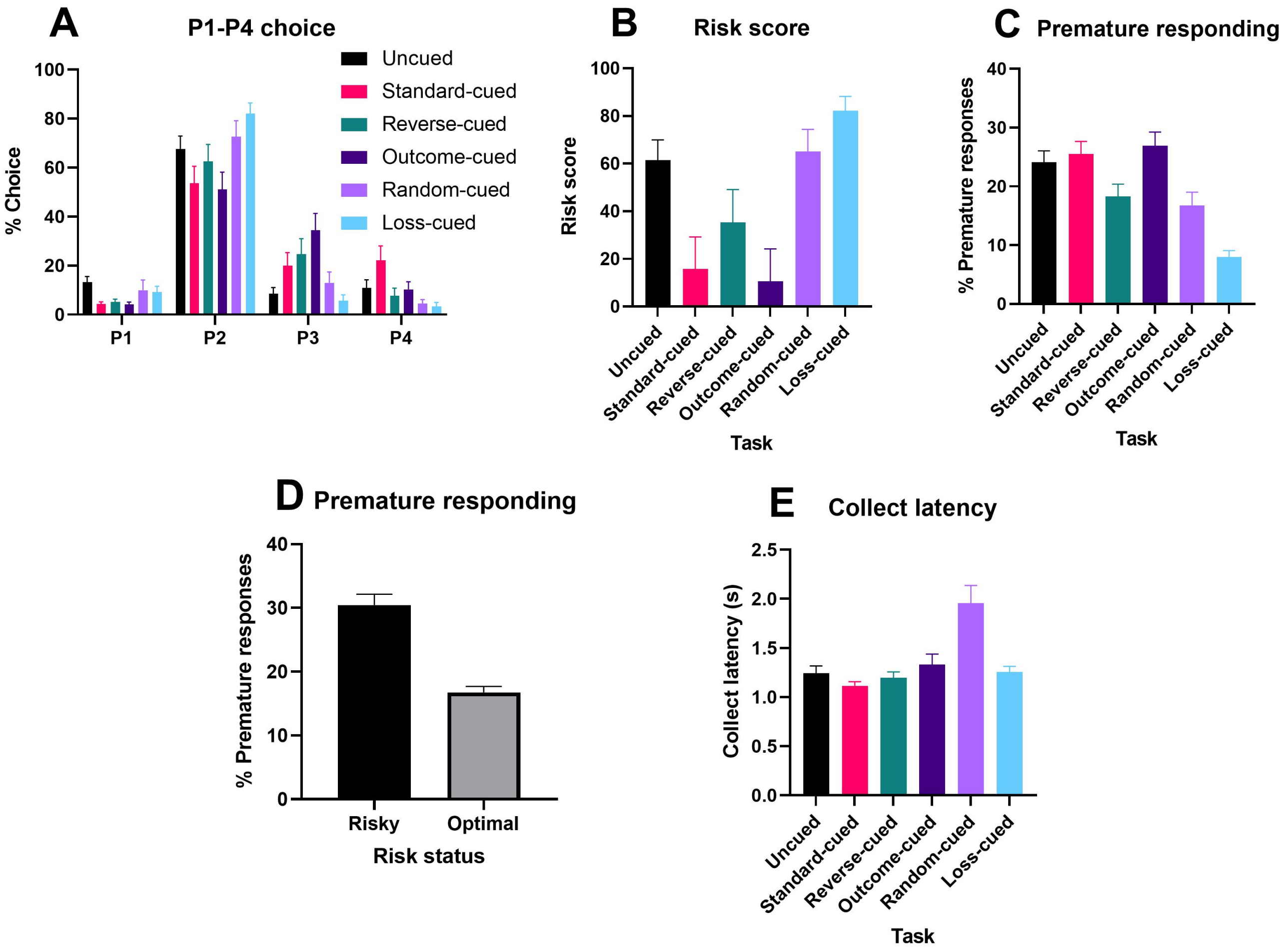


Fig. 2: Differences in baseline performance between task variants. Rats trained on tasks featuring win-paired cues exhibited higher levels of risk preference (increased P3/P4, decreased P1/P2 (A); lower risk score (B)). Premature responding (C) was lowest for the reverse-cued, random-cued, and loss-cued tasks. Across all tasks, risk-preferring rats had higher rates of premature responding than optimal rats (D). Latency to collect reward (E) was higher for the random-cued task than all other tasks. Data are expressed as mean + SEM.

Table 1: P1-P4 choice comparisons
Tukey HSD

	Task comparison		Mean difference	Significance
P1	Uncued	Standard	**.17**	0.007
		Reverse	0.14	0.06
		Outcome	**.18**	0.006
		Random	0.09	0.45
		Loss	0.08	0.62
	Standard	Reverse	-0.03	1.00
		Outcome	0.01	1.000
		Random	-0.08	0.62
		Loss	-0.09	0.45
	Reverse	Outcome	0.04	0.98

		Random	-0.05	0.93
		Loss	-0.06	0.83
	Outcome	Random	-0.09	0.54
		Loss	-0.10	0.38
	Random	Loss	-0.01	1.000
P2	Uncued	Standard	0.20	0.52
		Reverse	0.06	1.00
		Outcome	0.21	0.48
		Random	-0.07	0.99
		Loss	-0.20	0.52
	Standard	Reverse	-0.14	0.85
		Outcome	0.01	1.000
		Random	-0.27	0.20
		Loss	**-.40**	0.01
	Reverse	Outcome	0.15	0.81
		Random	-0.13	0.89
		Loss	-0.26	0.28
	Outcome	Random	-0.28	0.19
		Loss	**-.41**	0.01
	Random	Loss	-0.13	0.90
P3	Uncued	Standard	-0.16	0.47
		Reverse	-0.20	0.26
		Outcome	**-.35**	0.004
		Random	-0.04	0.999
		Loss	0.08	0.95
	Standard	Reverse	-0.04	0.998
		Outcome	-0.19	0.35
		Random	0.13	0.76
		Loss	0.25	0.10
	Reverse	Outcome	-0.15	0.65
		Random	0.17	0.53
		Loss	**.29**	0.04
	Outcome	Random	**.32**	0.02
		Loss	**.44**	0.0002
	Random	Loss	0.12	0.82
P4	Uncued	Standard	-0.14	0.37
		Reverse	0.06	0.98
		Outcome	0.02	1.000
		Random	0.11	0.72
		Loss	0.14	0.45
	Standard	Reverse	0.20	0.09
		Outcome	0.16	0.28
		Random	**.25**	0.01
		Loss	**.28**	0.004

Reverse	Outcome	-0.04	0.996
	Random	0.05	0.99
	Loss	0.08	0.90
Outcome	Random	0.09	0.86
	Loss	0.12	0.63
Random	Loss	0.03	0.999

Table 1: Comparisons of P1-P4 between task variants using Tukey's honest significant differences (HSD) test. Bolded values indicate a significant difference.

As is typical for analysis of data from this task and the IGT, an overall risk score was calculated by subtracting the percent choice of optimal options from the percent choice of the risky options ([P1 + P2] – [P3 + P4]). Animals with a risk score above zero were designated as "optimal", whereas rats with negative risk scores were classified as "risk-preferring". Average levels of risk score at the end of training differed significantly between tasks ($F(5, 170) = 6.62$, $p < .0001$, see Figure 2B). Task comparisons are reported in Table 2. In general, rats trained on tasks featuring win-paired cues were riskier than rats performing the uncued task. Data from the random-cued task did not differ significantly from the uncued task. Rats that learned the loss-cued task exhibited the lowest level of risk preference among all tasks.

Table 2: Risk score comparisons
Tukey HSD

Task comparison		Mean difference	Significance
Uncued	Standard	**45.51**	<.0001
	Reverse	**25.32**	.006
	Outcome	**51.75**	<.0001
	Random	-3.29	.99
	Loss	**-22.23**	.03
Standard	Reverse	-20.19	0.06
	Outcome	6.24	0.95
	Random	**-48.81**	<.0001
	Loss	**-67.75**	<.0001
Reverse	Outcome	**26.43**	0.006
	Random	**-28.62**	0.002
	Loss	**-47.56**	<.0001
Outcome	Random	**-55.05**	<.0001
	Loss	**-73.99**	<.0001
Random	Loss	-18.94	0.11

Table 2: Comparisons of risk score between task variants using Tukey's HSD test. Bolded values indicate a significant difference.

Premature responding

We next tested whether rats trained on each task variant differed in their level of motor impulsivity. This was measured by the proportion of premature responses made during the 5-second intertrial interval out of total trials. A significant difference was observed between tasks that was not dependent on risk status ($F(5,164) = 5.48$, $p = .0001$; see Figure 2C). Results of the *post-hoc* multiple comparisons can be found in Table 3. Rats trained on the loss-cued task had the lowest rate of premature responding compared to all other tasks. Reverse-cued rats and random-cued rats also exhibited a lower level of premature responding compared to the uncued, standard-cued, and outcome-cued rats.

Table 3: Premature responding comparisons
Tukey HSD

Task comparison		**Mean difference**	**Significance**
Uncued	Standard	-0.02	1.00
	Reverse	0.08	0.16
	Outcome	-0.03	0.96
	Random	**.10**	0.02
	Loss	**.24**	<0.0001
Standard	Reverse	**.095**	0.05
	Outcome	-0.01	1.00
	Random	**.12**	0.005
	Loss	**.25**	<0.0001
Reverse	Outcome	**-.11**	0.02
	Random	0.02	0.98
	Loss	**.16**	<0.0001
Outcome	Random	**.13**	0.002
	Loss	**.27**	<0.0001
Random	Loss	**.13**	0.001

Table 3: Comparisons of premature responding between task variants using Tukey's HSD test. Bolded values indicate a significant difference.

Across all task groups, risk-preferring rats had a significantly higher proportion of premature responses that optimal rats ($F(1,164) = 23.41$, $p < .0001$; see Figure 2D).

Other variables

Rats differed in their latency to collect reward across the task variants ($F(5,151) = 2.47$, $p = .04$; see Figure 2D). Results from the *post-hoc* multiple comparisons are displayed in Table 4, showing that rats trained on the random-cued task were significantly slower to collect reward than all other rats.

Table 4: Collect latency comparisons
Tukey HSD

Task comparison		Mean difference	Significance
Uncued	Standard	0.13	0.93
	Reverse	0.05	1.00
	Outcome	-0.09	0.99
	Random	**-.72**	<0.0001
	Loss	-0.02	1.00
Standard	Reverse	-0.08	0.99
	Outcome	-0.21	0.61
	Random	**-.84**	<0.0001
	Loss	-0.14	0.90
Reverse	Outcome	-0.13	0.93
	Random	**-.76**	<0.0001
	Loss	-0.06	1.00
Outcome	Random	**-.63**	0.0002
	Loss	0.07	1.00
Random	Loss	**.70**	<0.0001

Table 4: Comparisons of collect latency between task variants using Tukey's HSD test. Bolded values indicate a significant difference.

No differences between task variants were observed in latency to choose an option, trials completed, or omissions. Across all tasks, risk-preferring rats completed significantly fewer trials that optimal rats ($F(1,151) = 99.28$, $p < .0001$), as expected given that they experienced a higher number of lengthy time-out penalties.

Reinforcer devaluation

Choice

To determine which task variants resulted in inflexible choice patterns, rats were subjected to a reinforcer devaluation test in which they received *ad libitum* access to sucrose pellets for 1 hour prior to task performance. Data from the devaluation test were then compared to a baseline session during which no experimental manipulation occurred. A significant devaluation x choice x task effect was observed that was dependent on risk status (devaluation x choice x task x risk status: *F*(15, 314) = .44, *p* = .002). This effect was marginally significant in risk-preferring rats (F(15,66) = 1.79, p = .06). Effects broken down by task for risk-preferring animals can be found in Table 5; risk-preferring rats on the uncued, random-cued, and loss-cued tasks were grouped together due to low *n* (1-3 per task). Among the risky rats, only those trained on tasks without win-paired cues exhibited changes in choice patterns following reinforcer devaluation. In Figure 3A, choice of the P1-P4 options in risk-preferring rats are depicted as a difference in % choice between baseline and devaluation sessions (baseline subtracted from devaluation) for each task variant. This was done to highlight shifts in choice separate from overall group differences in the selection of the different options.

While the effects of devaluation did not differ by task in optimal rats (*F*(13,226) = 0.95, *p* = .50), a marginally significant choice x devaluation effect was observed in these rats (*F*(3,255) = 2.62, *p* = .06; see Figure 3B), indicating that some degree of shifting occurred in optimal rats that was not influenced by the presence or absence of cues.

Table 5: Devaluation in risk-preferring rats: P1-P4 choice

Task	F value	Degrees of freedom	P value
Uncued/random/loss (*n*=7)	**4.17**	3,9	.04
Standard (*n* = 7)	0.14	3,15	.93
Reverse (*n* = 3)	2.98	3,6	.12
Outcome (*n* = 14)	1.47	3,36	.24

Table 5: choice x devaluation interactions for each task in risk-preferring rats. Bolded values indicate a significant difference.

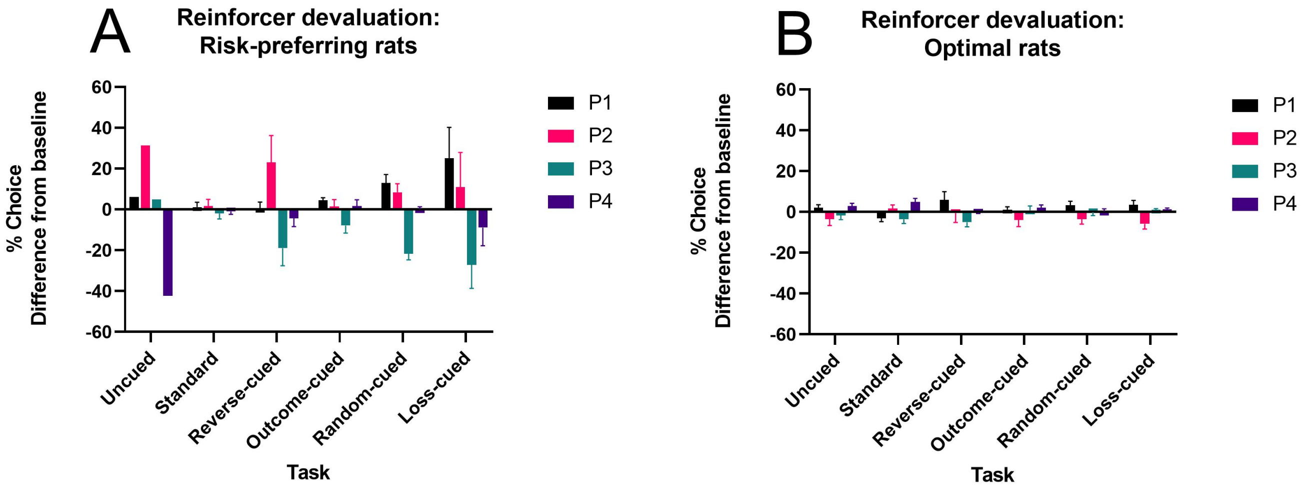


Fig. 3: Effects of sucrose pellet devaluation on choice preference. For risk-preferring rats (A), devaluation did not shift choice patterns in the tasks featuring consistent win-paired cues (standard, outcome-cued, reverse-cued). Choice patterns in optimal rats (B) shifted slightly, with no differences found between tasks. Data are expressed as the mean change in % choice from baseline + SEM to illustrate effects independent of differences in preference for each option between cohorts.

The observed shifts in P1-P4 choice resulted in a significant task-dependent shift in risk score in risk-preferring rats (devaluation x task: *F*(5,22) = 3.90, *p* = .01) but not optimal rats (*F*(5,85) = 1.53, *p* = .19). Results are summarized by task in Table 6. Similar to the choice results, only rats trained on tasks without win-paired cues exhibited shifts in risk preference following reinforcer devaluation.

Table 6: Devaluation in risk-preferring rats: Risk score

Task	F value	Degrees of freedom	P value
Uncued/random/loss (*n*=7)	**55.49**	1,3	.005
Standard (*n* = 7)	0.02	1,5	.89
Reverse (*n* = 3)	4.88	1,2	.16
Outcome (*n* = 14)	3.45	1,12	.09

Table 6: risk score x devaluation interactions for each task in risk-preferring rats. Bolded values indicate a significant difference.

Other variables

Latency to collect reward did not shift in response to devaluation ($F(1,107) = .55$, $p = .46$). Latency to choose an option significantly increased across all tasks ($F(1,107) = 71.38$, $p < .0001$), as did omissions ($F(1,107) = 9.75$, $p = .002$). Trials decreased in all rats, particularly optimal decision-makers (devaluation x risk status: $F(1,107) = 6.72$, $p = .01$; optimal: $F(1,85) = 80.66$, $p < .0001$; risky: $F(1,22) = 9.41$, $p = .006$). Premature responding significantly decreased across all groups ($F(1,107) = 63.32$, $p < .0001$).

Modeling learning dynamics of rGT cue variants

We investigated differences in the acquisition of each task variant by fitting several reinforcement learning models to completed trials in the first 5 sessions. Each of these models assumes that choice on every trial probabilistically follows latent Q-values for each option, and these are updated iteratively according to the experienced outcomes. Winning outcomes (R_{tr}) increase Q-values in a stepwise manner governed by the positive learning rate (η^{+}), according to a delta-rule update:

$$Q_x^{new} = Q_x^{old} + \eta^{+}(R_{tr} - Q_x^{old})$$

Three different models were designed to determine how losing outcomes decreased Q-values. Each model tests a different hypothesis as to how time-out penalties (T_{tr}) are transformed into an equivalent "cost" in sucrose pellets. The three models are summarized in Table 7.

Model name	Parameters	Punishment update
Scaled	4	$Q_x^{new} = Q_x^{old} + \eta^{-}(mT_{tr} - Q_x^{old})$
Scaled + offset	5	$Q_x^{new} = Q_x^{old} + \eta^{-}(b + mT_{tr} - Q_x^{old})$
Independent	7	$Q_x^{new} = Q_x^{old} + \eta^{-}(\omega_x T_{tr} - Q_x^{old})$

The scaled model assumes a linear relationship between the experienced time-out penalty durations, controlled by parameter m. The scaled + offset model features an additional offset parameter

b, allowing for a global increase or decrease in the impact of time-out penalties. Lastly, the independent model does not assume a linear relationship between durations; the time-out penalties for each option are transformed into an equivalent cost in pellets independently from each other ($\omega_1, \omega_2, \omega_3, \omega_4$ for P1-P4 respectively). The Q-values were then multiplied by the β parameter, which controls how closely rats' choices follow their latent Q-values (lower β value indicates more random choice across the four options). Parameters were estimated with Stan for both the group-level distributions and individual subjects using a hierarchical model structure (Carpenter et al., 2017).

To determine the best-fitting model, the Watanabe–Akaike information criterion (WAIC; Watanabe 2010) was calculated. This term assesses model fit whilst also penalizing for the number of parameters. A lower WAIC value indicates a better explanation of the data. Among the models tested, the independent RL model fit best to the data for each of the task groups (Fig 4; ΔWAIC >0 for all models compared to the independent punishment model). Thus, a non-linear transform for punishment duration best captures the choice data for all tasks. Additionally, the scaled + offset model fit the data better than the simplest scaled model – indicating that increased complexity of the model improved goodness-of-fit despite being penalized for a higher number of parameters.

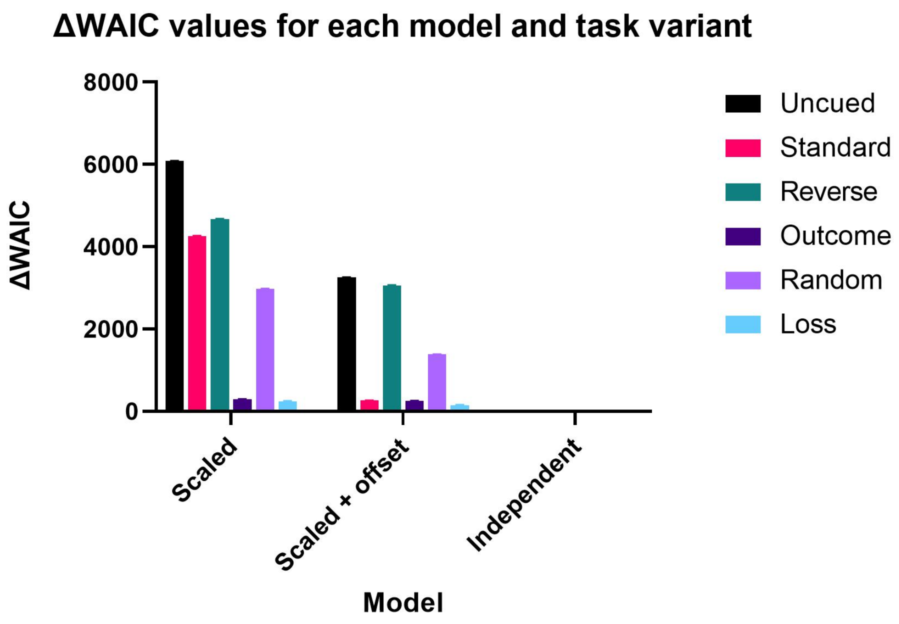

ΔWAIC values for each model and task variant

Fig. 4: difference in WAIC between each model and the independent model. Lower WAIC indicates a better explanation of the data. Error bars are SEM.

To confirm that the best-fitting independent punishment model captured the dominant features of the behavioural data, we simulated the probability of each option on each trial for 20 sessions, using the subject-level model parameter estimates (Fig. 5A). We then calculated the risk score for sessions 18-20 of the simulated data. Fewer significant differences were observed in the simulated data compared to the actual data (statistical tests available in supplemental table 1). Nevertheless, the overall pattern of results was preserved, with tasks featuring win-paired cues exhibiting lower risk scores than the uncued and loss-cued task, and the loss-cued task having the highest score. However, the random-cued task exhibited the lowest score, which was not observed in the actual data. This, in conjunction with the highly diffuse parameter estimates for the random-cued task, indicated that there was a degree of non-identifiability for this task and model. Thus, the scaled + offset model was also examined (Fig. 5B; statistical tests available in supplemental table 2). Data simulation from this model did not capture the reduction in risk score to the same degree for the win-paired cued tasks, nor is the elevated risk score for the loss-cued task present; however, the random-cued task is similar in score to the uncued task, as observed in the real data. In addition, the uncued task is significantly different from the standard-cued and outcome-cued task, which is not present in the independent model data simulation. This may be due to the diffuse group-level parameter estimates that were also observed for the uncued task, indicating some level of difficulty in identifying precise estimates. Accordingly, the group-level parameter estimates for both the independent and scaled + offset models are examined below. Group-level parameter estimates for the scaled model are depicted in supplemental figure 1.

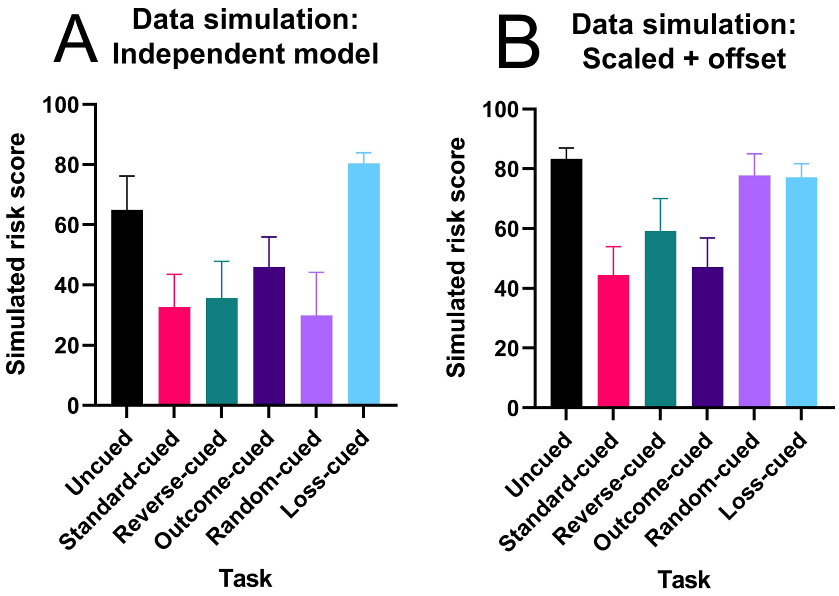


Fig. 5: Average risk score (sessions 18-20) for the independent (A) and scaled + offset (B) models simulated with the subject-level parameter estimates for each task variant.

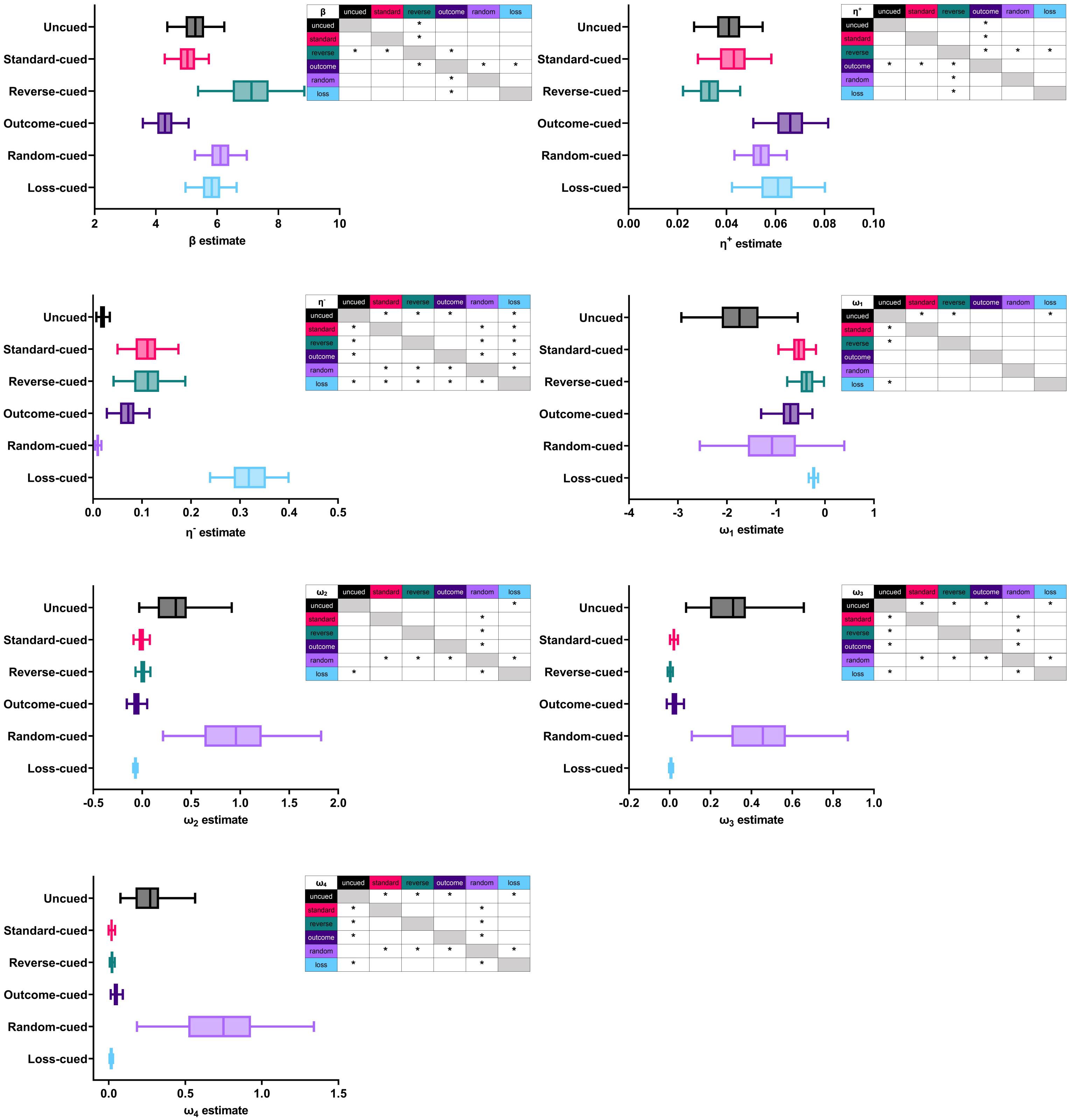
Group-level differences in learning

To identify differences in learning dynamics between tasks, the 95% highest density intervals (HDI) for group-level mean parameter estimates were compared between tasks. Differences were considered credible when the 95% HDI for the sample difference between two mean estimates did not include zero. As the independent model fit the data from all the tasks the best, we first examined the mean estimates for each parameter from these model fits.

Independent model.

The posterior distributions for the group-level parameter estimates for each task are displayed in Figure 6, with credible differences indicated by asterisks within the inset tables. Notably, the mean beta parameter estimate for the outcome task is lower than all except the standard task, indicating that choice patterns on this task did not follow latent Q-values as closely. The outcome-cued task also exhibited a higher positive learning rate than many other tasks, whereas the estimate for the reverse-cued task is lower than the overall average. Also of note is the high negative learning rate for the loss-cued task.

The 95% HDIs for the P1-P4 weights are considerably more diffuse for the uncued and random-cued tasks compared to the others, as well as an unusually low mean estimate for the negative learning rate. As previously mentioned, data simulation for these tasks did not completely recapitulate the observed behavioural data. Therefore, the scaled + offset model was also examined, as the reduction in model complexity resulted in an increase in precision for uncued and random-cued task parameter estimates.



β	uncued	standard	reverse	outcome	random	loss
uncued			*			
standard			*			
reverse	*	*		*		
outcome			*		*	*
random				*		
loss				*		

η^+	uncued	standard	reverse	outcome	random	loss
uncued				*		
standard				*		
reverse				*	*	*
outcome	*	*	*			
random			*			
loss			*			

η^-	uncued	standard	reverse	outcome	random	loss
uncued		*	*	*		*
standard	*				*	*
reverse	*				*	*
outcome	*				*	*
random		*	*	*		*
loss	*	*	*	*	*	

ω_1	uncued	standard	reverse	outcome	random	loss
uncued		*	*			*
standard	*					
reverse	*					
outcome						
random						
loss	*					

ω_2	uncued	standard	reverse	outcome	random	loss
uncued						*
standard					*	
reverse					*	
outcome					*	
random		*	*	*		*
loss	*				*	

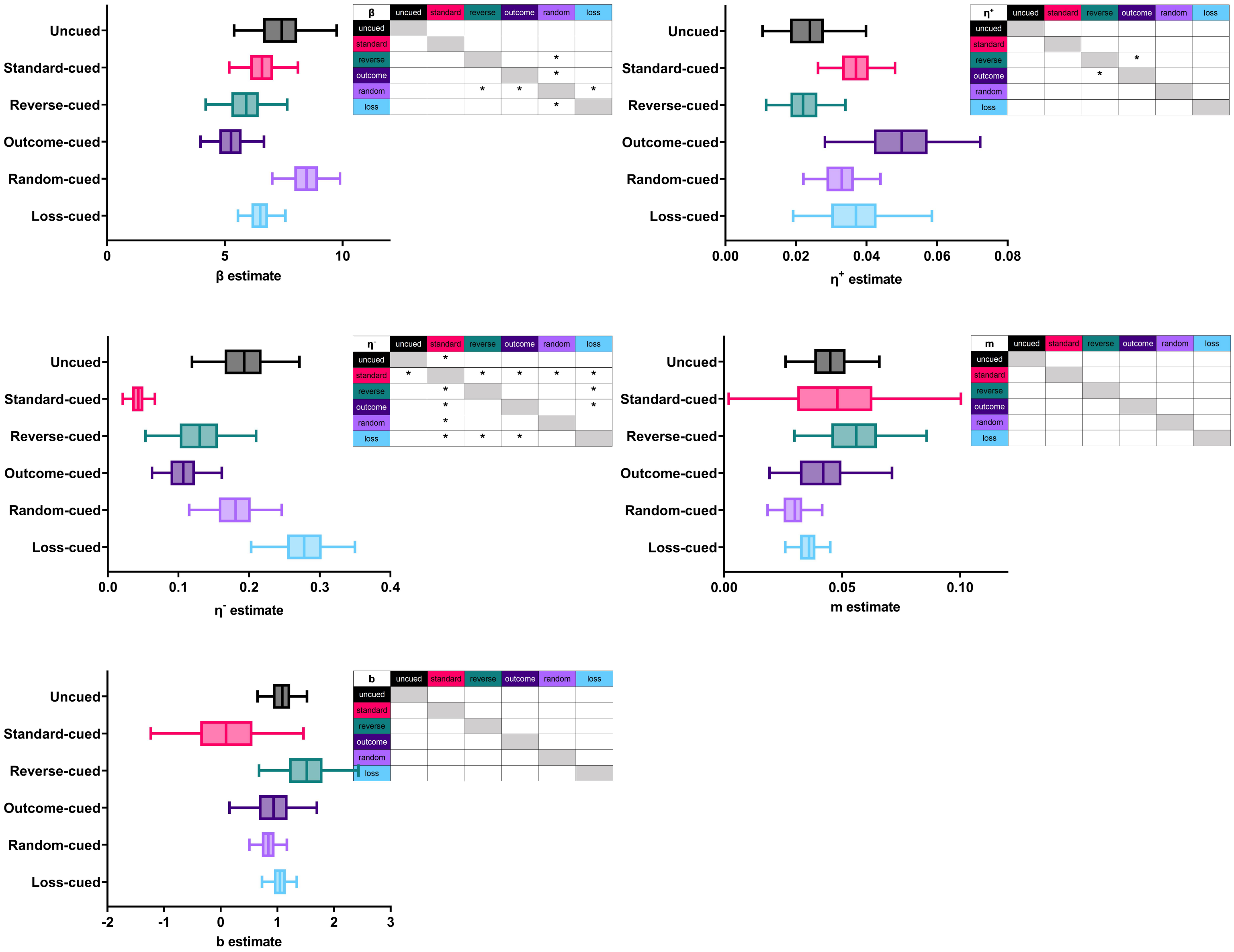
ω_3	uncued	standard	reverse	outcome	random	loss
uncued		*	*	*		*
standard	*				*	
reverse	*				*	
outcome	*				*	
random		*	*	*		*
loss	*				*	

ω_4	uncued	standard	reverse	outcome	random	loss
uncued		*	*	*		*
standard	*				*	
reverse	*				*	
outcome	*				*	
random		*	*	*		*
loss	*				*	

Fig. 6: Group-level posterior estimates of independent model parameters. Asterisks within the inset tables mark parameters for which the 95% HDI of the sample difference did not contain zero, indicating a credible difference. For each distribution, the box demarcates the interquartile interval and the whiskers demarcate the 95% HDI.

Scaled+offset model.

The pattern of results observed for the scaled + offset model (Fig. 7) are largely similar to the independent model, with a few notable exceptions. Here, the random task has a higher mean beta estimate than the reverse-, outcome-, and loss-cued tasks. The positive learning rate for the outcome task is higher than the overall average, although it is only credibly different from the reverse-cued task. Additionally, the standard-cued task mean estimate for the negative learning rate is lower than all other tasks. Lastly, as previously observed, the loss-cued task has the highest mean estimate for negative learning rate, but not all task comparisons are credibly different.



β	uncued	standard	reverse	outcome	random	loss
uncued						
standard						
reverse					*	
outcome					*	
random			*	*		*
loss					*	

η^+	uncued	standard	reverse	outcome	random	loss
uncued						
standard						
reverse				*		
outcome			*			
random						
loss						

η^-	uncued	standard	reverse	outcome	random	loss
uncued		*				
standard	*		*	*	*	*
reverse		*				*
outcome		*				*
random		*				
loss		*	*	*		

m	uncued	standard	reverse	outcome	random	loss
uncued						
standard						
reverse						
outcome						
random						
loss						

b	uncued	standard	reverse	outcome	random	loss
uncued						
standard						
reverse						
outcome						
random						
loss						

Fig. 7: Group-level posterior estimates of scaled + offset model parameters. Asterisks within the inset tables mark parameters for which the 95% HDI of the sample difference did not contain zero, indicating a credible difference. For each distribution, the box demarcates the interquartile interval and the whiskers demarcate the 95% HDI.

Parameters predicting risk preference on the rGT

We next tested whether any of the subject-level parameter estimates in the independent or scaled + offset model could reliably predict risk preference scores at the end of training. For the independent model, we found that the parameters controlling the weight of the time-out penalties specifically for the risky options (P3 weights: $R^2 = .056$, $F(1,163) = 9.60$, $p = .002$, Fig. 8A; P4 weights: $R^2 = .026$, $F(1,163) = 4.35$, $p = .04$, Fig. 8B) were predictive of rats' ultimate risk preferences. The positive learning rate reached marginal significance ($R^2 = .021$, $F(1,163) = 3.53$, $p = .06$). These predictive relationships indicate that risk-preferring rats experienced significantly lower time-out penalty costs for the risky options compared to optimal rats.

For the scaled + offset model, the negative learning rate parameter estimates predicted risk score at the end of training ($R^2 = .029$, $F(1,163) = 4.80$, $p = .03$, Fig. 8B). This indicates that a lower negative learning rate was associated with greater risk preference.

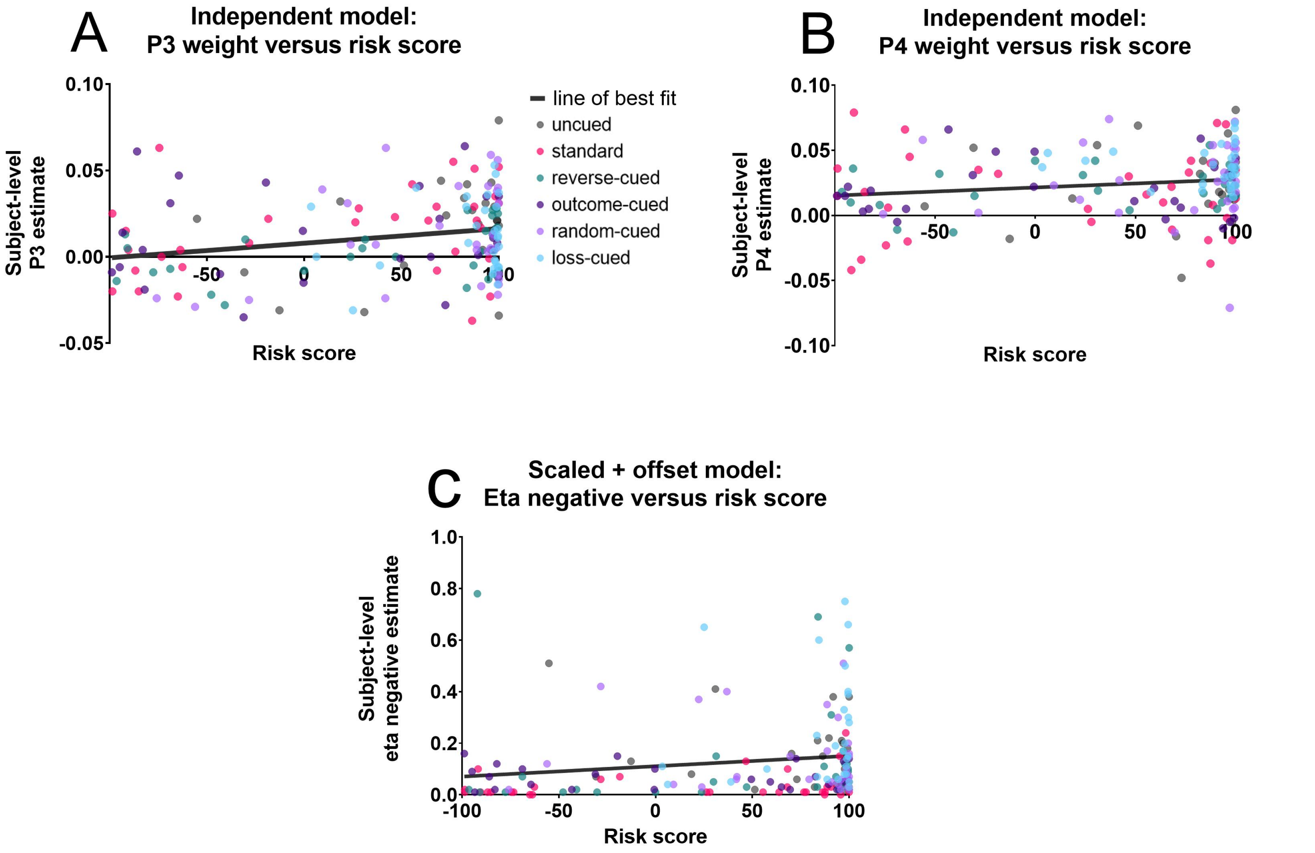


Fig. 8: subject-level parameter estimates for the risky option weights from the independent model (A,B) and the negative learning rate from the scaled + offset model (C) predicted risk preference at the end of training.

Discussion

Here, we showed that audiovisual cues drive risky choice on the rat gambling task (rGT) only if they are reliably, but not exclusively, win-paired. This was demonstrated by higher levels of risky choice in rats trained on the standard-cued, reverse-cued, and outcome-cued variants of the rGT. Computational analysis of the acquisition phase using reinforcement learning models revealed that differences in decision making were largely captured by parameters that control learning from punishments. These parameters predicted risk score at the end of training, indicating that risk-preferring rats discounted losses to a greater degree than optimal rats. There was also some evidence of increased learning from rewarded trials, particularly when all outcomes were cued.

These results largely confirm and build upon the previous report investigating learning dynamics of the cued versus uncued rGT (Langdon et al., 2019). We previously observed that the addition of reward-paired cues to the task resulted in insensitivity to punishments, particularly on the risky options, and that time-out penalty weights could predict risk score at the end of training. The current studies extend these results to suggest that this relationship can be bidirectionally modulated, as loss-paired cues reduce risky choice and increase learning from losing outcomes. Furthermore, pairing cues with wins seems to dominate the decision-making process. Cuing the losses when the wins are also cued has no risk-reducing effect, and in fact may even further potentiate risky choice.

While the corroborating results across models lends strength to these conclusions, the independent model did not precisely capture the learning dynamics for some groups, despite being the best fitting for all tasks. In particular, the uncued and random-cued tasks had diffuse group-level estimates for some parameters. The unusually peaked distribution near zero for the negative learning rate and the diffuse distributions for the P1-P4 weights suggest that there was difficulty in isolating the influence of these parameters. This may be because, given the low number of risk-preferring rats on these tasks, there were relatively few of certain trial types (e.g., losses on P3/P4) and the more complex model could not differentiate between the speed versus the size of the update to the latent values following positive and negative outcomes.

Alternatively, rats on the uncued- and random-cued task may have had a more variable choice pattern early in training because of the nature of the task. Adding predictable cues to the task may encourage consistency in the development of choice patterns, albeit in opposite directions depending on the cued outcome; win-paired cues elicit risky choice, and loss-paired cues elicit risk aversion. Conversely, rats trained on tasks that lack salient cues or with cues that do not carry useful information about outcomes may not be as stereotyped in their behaviour early in training. For example, they could switch strategies as sessions progress, or there may be more subjects who maintain a high degree of variability across sessions. Therefore, cues (provided they have an interpretable relationship with outcomes) may drive consistency and constrain the way in which the task is interpreted. Developing models in which parameter estimates can evolve over time may allow for potential differences in decision-making variability to come to light.

Results from the reinforcer devaluation test provide additional support for differences in decision-making processes when win-paired cues are present, in that risk-preferring rats trained on these tasks were not sensitive to changes in reinforcer value. This indicates that such cues can render choice patterns inflexible. However, no differences were found between tasks for optimal rats, and they were overall less sensitive to changes in reinforcer value. Optimal rats may therefore be indifferent to fluctuations in reward value such as occasional reward omission on the low-risk options, or in this case, reinforcer devaluation. Additionally, if we assume that frequency of winning probability becomes more desirable than reward size, optimal rats can only move to P1 to maximize win frequency, whereas risk-preferring rats have more options to which they may shift. Nevertheless, the fact that risk-preferring rats trained on these tasks did not shift suggests that win-paired cues, particularly when they track reward size, inhibit flexible responding.

While differences in reinforcer devaluation tests in risky versus optimal rats have not been previously observed on the rGT, the cohort sizes of the present study far exceed previous reports, which may have been underpowered to detect such differences. The results from risk-preferring rats corroborate previous studies demonstrating that rats trained on the uncued task are sensitive to this manipulation, whereas rats trained on the cued task are not (Zeeb & Winstanley, 2013; Hathaway et al., 2021). This could be due to either enhanced habit formation or hypoactive, or otherwise maladaptive, goal-directed control. Altering serotonergic signaling in the lateral orbitofrontal cortex (OFC) can restore sensitivity to reinforcer devaluation in rats trained on the cued rGT, indicating that prefrontal cortices, and presumably impaired goal-directed control, play a role in inflexibility induced by reward-paired cues (Hathaway et al., 2021). Indeed, given the role of the lateral OFC in updating stored action-outcome contingencies, cue-guided learning, and in the acquisition of uncued rGT (Izquierdo, 2017; Amodeo et al., 2017, Zeeb & Winstanley, 2011), this region could be mediating both the differential processing of rewards and punishments across different variants of the rGT and cue-induced inflexibility. Alternatively, the task bracketing hypothesis from Vandaele et al. (2017) posits that the inclusion of salient cues in decision-making tasks (lever insertion, in their case) encourages the formation of rigid stimulus-response patterns, rather than generally reducing cognitive flexibility. Future studies could test other facets of cognitive flexibility both during and outside of the rGT (e.g., extinction training on the rGT, or probabilistic reversal learning after rGT training) to further explore these lines of thought.

In addition to inducing inflexibility, win-paired cues on the outcome-cued and standard-cued tasks impacted choice patterns in a relatively comparable manner. These results disprove the “state” hypothesis described in the introduction, in which we suggested that increasing the similarity of winning and losing trials may permit better integration of the time-out penalties into the stored values of each option. We instead observed that outcome-cued rats were equally risky as those trained on the standard-cued paradigm. Cuing losses did not increase learning about those trial types; instead, it may have disguised them as wins. Losses disguised as wins (LDWs) are a feature of modern multi-line slot machines in which win-related cues are played when a small payout that is less than the bid is won, leading the player to believe they've won money when in fact they have experienced a net loss. These LDWs can therefore be miscategorized as wins (Dixon et al., 2010; Jensen et al., 2013). Marshall and Kirpatrick (2017) applied a reinforcement learning model to behavioural data from their task investigating the LDW effect in rodents and showed that playing win-related sensory feedback during losses elevated stay biases on the high-risk option by increasing its value. A similar mechanism may be at play in the outcome-cued task. Conversely, cuing only losses may oppose the LDW effect, as others have shown that playing loss-associated cues during LDWs can permit subjects to correctly categorize them as losses (Dixon et al., 2015). It is interesting to note that rats on the outcome-cued task had the lowest beta parameter estimate. This may suggest the model did not capture the development of their choice patterns to the same degree as the other tasks. It could be that adding a stay-bias parameter similar to the model by Marshall and Kirkpatrick (2017) would better encapsulate the learning dynamics for rats on this task.

An alternate hypothesis for the impact of win-paired cues on decision making comes from research investigating the role of dopamine in the perception of time. Recent evidence suggests that the timing of dopamine signals influences whether subjective time speeds up or slows down (Jakob et al., 2021). Hence, it could be that cued rewards alter the subjective experience of the time-out penalty duration via a dopaminergic mechanism. Indeed, the standard-cued task is more sensitive to dopamine manipulations (Barrus & Winstanley, 2016). Dopamine signals provoked by win-paired cues may reduce the experienced duration of the time-out penalties such that their impact on the latent value of each option is diminished. Measurements of dopamine signals on-task using fiber photometry could be incorporated into a model to test this hypothesis (see Jakob et al., 2021).

While pairing cues with wins is sufficient to drive risky choice, cue complexity and magnitude appear to play a minor role, as rats on the reverse-cued task were significantly less risky than the outcome-cued rats, and marginally different from the standard-cued rats. Additionally, parameter estimates for rats trained on the reverse-cued task did not always align with the other tasks featuring win-paired cues (e.g., lower learning rate for rewarded trials). These rats also exhibited a lower rate of premature responding compared to all other tasks except for the loss-cued task. This may indicate that matching cue size and complexity to reward size can potentiate motor impulsivity. Indeed, when the salience of reward-predictive cues matches the size of the reward, activity in the nucleus accumbens is amplified (Knutson et al., 2001), which may be diminished when cue size inversely scales with reward. As activity within the nucleus accumbens is critically involved in motor impulsivity on similar behavioural tasks (Economidou et al., 2012; Pattij et al., 2007), reduction of this signal could explain the low rate of premature responding in these rats.

Consistently pairing cues with wins proved to be a necessary component to induce risky choice, as playing cues randomly on 50% of trials regardless of outcome did not significantly shift risk preference compared to the uncued variant. We originally thought that the increased sensory stimulation from the random cues could increase arousal and therefore risk preference. However, rats instead learned to disregard these cues and were perhaps less engaged in the task, as indicated by the longer latencies to collect reward and reduced levels of premature responding. That being said, this finding does not disprove the hypothesis that increased arousal leads to riskier choice patterns; it may be that the cue-reward relationship increases arousal in a way that random cues cannot. Recent results implicating norepinephrine in cue-induced risky choice would suggest that arousal may contribute to the impact of cues on decision making (Chernoff et al., 2021). It would be interesting to pretrain rats on the association between the cues and reward prior to rGT training on the random-cued task; in that case, increased risk preference may be observed. This may represent an intriguing model of the effect of ambient lights and sounds of a casino on gambling behaviour.

Pairing cues with losses would also ostensibly increase arousal, however their behavioural impact was quite distinct from that of win-paired cues. Indeed, rats trained on the loss-cued task were the least risk-preferring out of all the groups, including the uncued task. This would suggest that, while the uncued task is usually regarded as a control for the cued task(s), it may also be a deviation from how optimal rats can be. Indeed, Langdon et al. (2019) found that all rats across both the uncued and standard-cued task were globally less sensitive to the time-out penalties, and that differences in risk preference arose from the degree of this reduced sensitivity. Conversely, rats on the loss-cued task appear to be more sensitive to losses. Thus, they could be more proactively risk-avoiding, and rats on the uncued task may be more willing to sample from the risky options despite having an overall optimal decision-making profile.

Overall, the results from these studies indicate that outcome-associated cues play a significant role in decision-making processes, and their effect is highly dependent on the outcome type with which they are associated. Differences in choice patterns are largely a result of changes to the relative impact of losses on decision making, as revealed by the effect of different cue paradigms on group-level parameter estimates capturing learning from losses in the tested reinforcement learning models. This provides critical insight into the influence of the rich sensory environment in casinos and other forms of gambling, particularly the addictive allure of electronic gaming machines. Furthermore, these analyses demonstrate the power of combining modeling approaches with careful behavioural manipulations to inform our understanding of action selection in complex decision-making scenarios.

Methods

Subjects

Subjects were four cohorts of 32-64 male Long Evans rats (Charles River Laboratories, St Constant, QC, Canada) weighing 275–300 g upon arrival to the facility. One to two weeks following arrival, rats were food-restricted to 14 g of rat chow per day and were maintained at least 85% body weight of an age- and sex-matched control. Water was available *ad libitum*. All subjects were pair-housed or trio-housed in a climate-controlled colony room under a 12 h reverse light-dark cycle (21 °C; lights off at 8am). Huts and paper towel were provided as environmental enrichment. Behavioural testing took place 5 days per week. Housing and testing conditions were in accordance with the Canadian Council of Animal Care, and experimental protocols were approved by the UBC Animal Care Committee.

Behavioural apparatus

Testing took place in 32 standard five-hole operant chambers, each of which was enclosed in a ventilated, sound-attenuating chamber (Med Associates Inc, Vermont). Chambers were fitted with an array composed of five equidistantly spaced response holes. A stimulus light was located at the back of each hole, and nose-poke responses into these apertures were detected by vertical infrared beams. On the opposite wall, sucrose pellets (45 mg; Bioserv, New Jersey) were delivered to the magazine via an external pellet dispenser. The food magazine was also fitted with a tray light and infrared sensors to detect sucrose pellet collection. A house light could illuminate the chamber. The operant chambers were operated by software written in Med-PC by CAW, running on an IBM-compatible computer.

Behavioural testing

Rats were first habituated to the operant chambers in two daily 30-minute sessions, during which sucrose pellets were present in the nose-poke apertures and food magazine. Rats were then trained on a variant of the five-choice serial reaction time task and the forced-choice variant of the rGT, as described in previous reports (Zeeb, Robbins, & Winstanley, 2009; Barrus & Winstanley, 2016).

A task schematic of the rGT is provided in Figure 1. During the 30-minute session, trials were initiated by making a nose-poke response within the illuminated food magazine. This response extinguished the light, which was followed by a five-second inter-trial interval (ITI) in which rats were required to inhibit their responses to proceed with the trial. Any response in the five-hole array during the ITI was recorded as a premature response and punished by a five-second time-out period, during which the house light was illuminated and no response could be made. Following the ITI, apertures 1, 2, 4, and 5 in the five-hole array were illuminated for 10 seconds. A lack of response after 10 seconds was recorded as an omission, at which point the food magazine was re-illuminated and rats could initiate a new trial. A nose-poke response within one of the illuminated apertures was either rewarded or punished according to that aperture's reinforcement schedule. Probability of reward varied among options (0.9-0.4, P1-P4), as did reward size (1-4 sucrose pellets). Punishments were signalled by a light flashing at 0.5 Hz within the chosen aperture, signalling a time-out penalty which lasted for 5-40 seconds depending on the aperture selected. The task was designed such that the optimal strategy to earn the highest number of sucrose pellets during the 30-minute session would be to exclusively select the P2 option, due to the relatively high probability of reward (0.8) and short, infrequent time-out penalties (10 s, 0.2 probability). While options P3 and P4 provide higher per-trial gains of 3 or 4 sucrose pellets, the longer and more frequent time-out penalties associated with these options greatly reduces the occurrence of rewarded trials. Consistently selecting these options results in fewer sucrose pellets earned across the session and are therefore considered disadvantageous. The position of each option was counterbalanced across rats to mitigate potential side bias. Half the animals in each project were trained on version A (left to right arrangement: P1, P4, P2, P3) and the other half on version B (left to right arrangement: P4, P1, P3, P2).

Task variants

Six variants of the task were used in this experiment (*n* = 28-32 rats per task variant). On the uncued task, winning trials were signalled by the illumination of the food magazine alone. On the standard-cued task, reward delivery occurred concurrently with 2-second compound tone/light cues. Cue complexity and variability scaled with reward size, such that the P1 cue consisted of a single tone and illuminated aperture, and the P4 cue consisted of multiple tones and flashing aperture lights presented in four different patterns across rewarded trials. The reverse-cued task featured an inversion of the cue-reward size relationship, such that the longest and most complex cue occurred on P1 winning trials, and P4 winning trials were accompanied by a single tone and illuminated aperture. On the outcome-cued task, all trial outcomes were accompanied by an audiovisual cue (i.e., during reward delivery and at the onset of the time-out penalty). The random-cued task consisted of cues that occurred on 50% of trials, regardless of outcome. Lastly, on the loss-cued task, cues occurred only on losing trials, at the onset of the time-out penalty. Cue complexity and magnitude scaled with reward size/time-out penalty length for the outcome-, random-, and loss-cued variants of the task (i.e., same pattern as the standard-cued task).

Reinforcer devaluation

128 rats (*n* = 12-28 per task version) underwent a reinforcer devaluation procedure. This procedure took place across two days. On the first day, half of the rats were given *ad libitum* access to the sucrose pellets used as a reward on the rGT for 1 hour prior to task initiation. The remaining rats completed the rGT without prior access to sucrose pellets. Following a baseline session day for which no sucrose pellets were administered prior to the task to any rats, the groups were then reversed and the other half were given 1-hour access to sucrose pellets.

Behavioural measures and data analysis

All statistical analyses were completed using SPSS Statistics 27.0 software (SPSS/IBM, Chicago, IL, USA). As per previous reports, the following rGT variables were analyzed: percentage choice of each option (number of times option chosen/total number of choices × 100), risk score (calculated as percent choice of [(P1 + P2) − (P3 + P4)]), percentage of premature responses (number of premature responses/total number of trials initiated × 100), sum of omitted responses, sum of trials completed, and average latencies to choose an option and collect reward. Variables that were expressed as a percentage were subjected to an arcsine transformation to limit the effect of an artificially imposed ceiling (i.e., 100%). Animals with a mean positive baseline risk score were designated as “optimal”, whereas rats with negative risk scores were classified as “risk-preferring”.

For baseline analyses, mean values for each variable were calculated by averaging across four consecutive sessions that were deemed statistically stable (i.e., session and/or session x choice interaction were not significant in a repeated-measures ANOVA; following approximately 35-40 training sessions). Task (six levels: uncued, standard-cued, reverse-cued, outcome-cued, random-cued, loss-cued) and risk status (two levels: optimal, risk-preferring) were included as between-subjects factors for all baseline analyses. Choice data were analyzed with a two-way repeated measures ANOVA with choice (four levels: P1, P2, P3, and P4) as within-subject factors. For the analysis of the reinforcer devaluation data, devaluation (two levels: baseline, devaluation) and choice (four levels: P1-P4) were the within-subject factors and task version and risk status were the between-subjects factors.

For all analyses, if sphericity was violated as determined by Mauchley's test, a Huynh–Feldt correction was applied, and corrected *p* values' degrees of freedom were rounded to the nearest integer. Results were deemed to be significant if *p* values were less than or equal to an α of .05. Any main effects or interactions of significance were further analyzed via *post hoc* one-way ANOVA or Tukey’s tests. Any *p*-values > .05 but < .09 were reported as a statistical trend.

Hierarchical modeling of task learning

A full description of the modeling approach can be found in Langdon et al. (2019). Valid choice trials from the first 5 sessions were concatenated into one long session and trial-by-trial preferences were modeled using three reinforcement learning models (RL; Sutton & Barto, 1998). Each model was fit separately to each task variant group, thus allowing for the possibility that different RL models might perform better at predicting choice for each of the groups. Data from 11 rats were excluded due to missing sessions, experimenter error, or technical issues. This left a total of 24 rats in the uncued task group, 32 rats in the standard-cued task group, 25 rats in the outcome-cued task group, and 28 rats in the reverse-, random-, and loss-cued task groups.

Each of these models assumes that choice on every trial probabilistically follows latent *Q*-values for each option, and these are updated iteratively according to the experienced outcomes. For our models, the probability of choosing option P_x on each trial follows the learned *Q*-values for $x =$ [1,2,3,4] according to the softmax decision rule:

$$p(P_x) = \frac{e^{\beta Q_x}}{\sum_{y=1}^{4} e^{\beta Q_y}},$$

where $p(P_x)$ is the probability of choosing option P_x, Q_x is the learned latent value of option *x*, and β is the inverse temperature parameter that controls how strongly choice follows the latent *Q*-values rather than a random (uniform) distribution over the four options. In each learning model, we assume learning of latent *Q*-values from positive outcomes follows a simple delta-rule update:

$$Q_x^{new} = Q_x^{old} + \eta^{+}(R_{tr} - Q_x^{old}),$$

where $\eta+$ is a learning rate parameter that governs the step-size of the update, $R_{tr} > 0$ is the number of pellets delivered on a given winning trial, and Q_x is the latent value for the chosen option *x* on a given trial.

Q-values for learning from punishments were updated differently depending on the model. In each case, we sought to model the negative impact of time-out penalties on choice by transforming the duration of the penalty into an equivalent "cost" in sucrose pellets. Each model tests a different hypothesis on the transform of the punishments, with a separate negative learning rate η–. These are summarized in Table 1.

Model name	Parameters	Punishment update
Scaled	4	$Q_x^{new} = Q_x^{old} + \eta^-(mT_{tr} - Q_x^{old})$
Scaled + offset	5	$Q_x^{new} = Q_x^{old} + \eta^-(b + mT_{tr} - Q_x^{old})$
Independent	7	$Q_x^{new} = Q_x^{old} + \eta^-(\omega_x T_{tr} - Q_x^{old})$

In the scaled punishment model, we assume that the equivalent punishment for a time-out penalty on each losing trial scales linearly with the duration of the punishment. T_{tr}>0 is the time-out penalty duration in seconds on a given losing trial and *m* is a scaling parameter that maps time-out duration into an equivalent cost in pellets (i.e., has units pellets/s). The scaled + offset model is the same as the scaled punishment model but features an additional offset parameter *b*, which removes the constraint that the linear transform between time-out penalty duration and equivalent cost is zero for zero duration. Lastly, in the independent model, we model individual punishment weights for each outcome, allowing a nonlinear mapping between experienced duration and the equivalent cost in pellets on each trial. Equivalent costs for each option are controlled independently by ω_x for each option P_x.

For every model, *Q*-values were initialized at zero for the first session, and we assumed *Q*-values at the start of a subsequent session (on the next day for example) were the same as at the end of the previous session (i.e., we modeled no intersession effects on learning). Each model was fit to the entire set of choices for each group of rats using Hamiltonian Monte Carlo sampling with Stan to perform full Bayesian inference and return the posterior distribution of the model parameters conditional on the data and specification of the model (Carpenter et al., 2017). In each case, we partially pooled choice data across individual rats in a hierarchical model to simultaneously determine the distribution of individual- and group-level model parameters. We implemented a noncentered parameterization (a.k.a. the "Matt trick") for group-level β, η+, and η− in each model, as this has been shown to improve performance and reduce autocorrelation between these group-level parameters in hierarchical RL models (Ahn et al, 2017).

Each model was fit using four chains with 1000 or 2000 steps each (1000 burn-in), yielding a total of 4000 or 8000 posterior samples. To assess the convergence of the chains, we computed the $\hat{R}$ statistic (Gelman et al. 2013), which measures the degree of variation between chains relative to the variation within chains. Across all three models, no parameter had $\hat{R} > 1.1$, and the mode was 1.00, indicating that for each model all chains had converged successfully.

To measure the difference between group-level parameters, we used highest density intervals (HDI; Kruschke, 2014). The HDI is the interval which contains the required mass such that all points within the interval have a higher probability density than points outside the interval. Differences were considered credible when the 95% HDI for the sample difference between two mean estimates did not include zero. To compare the overall performance of each model, we computed the Watanabe–Akaike information criterion (WAIC; Watanabe, 2010), which, like AIC or BIC, provides a metric to compare different models fit to the same dataset. The WAIC is computed from the pointwise log-likelihood of the full posterior distribution (thereby assessing model fit) with a second term penalizing for model complexity.

Supplemental tables and figures

Table S1: Independent model simulated risk score comparisons
Tukey HSD

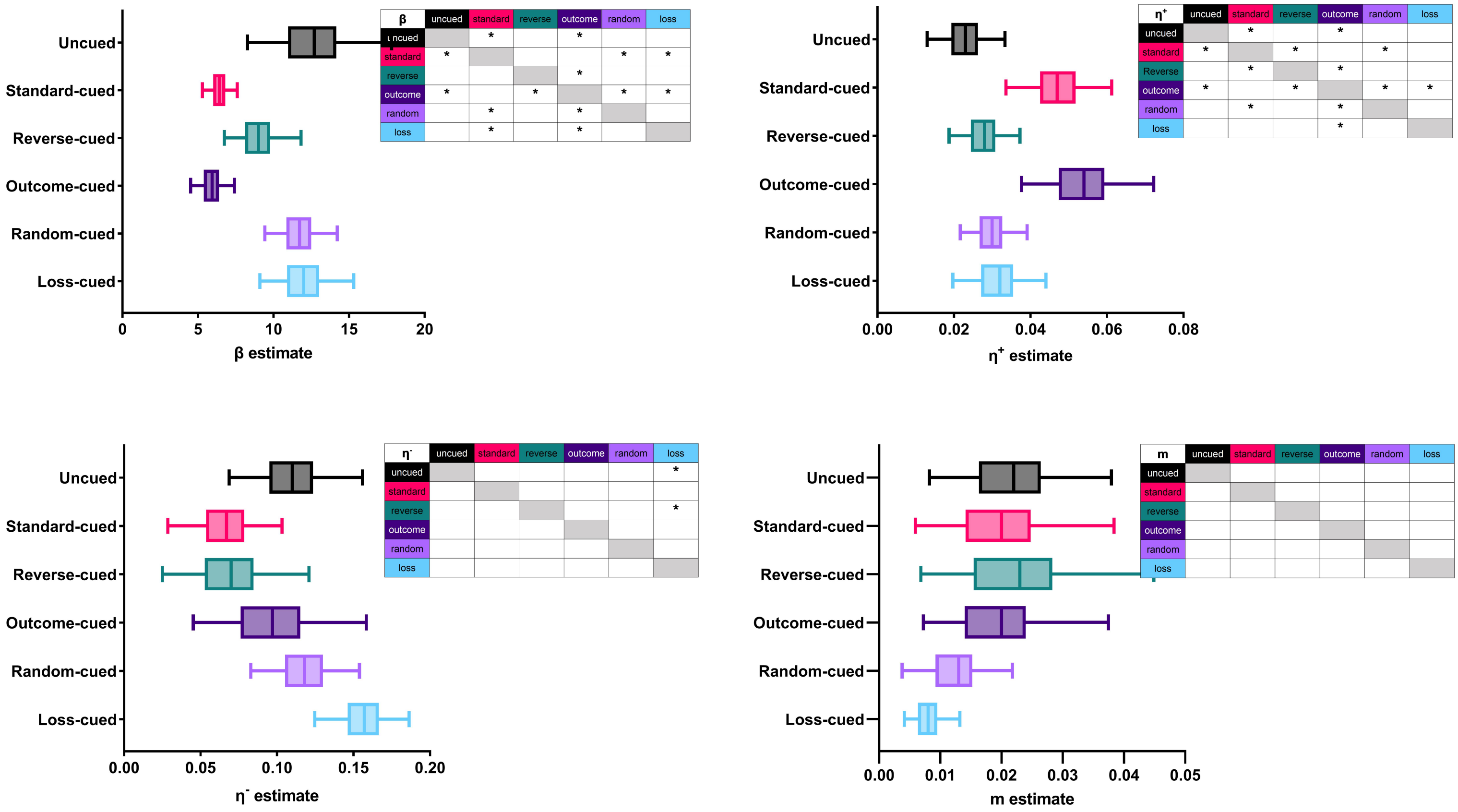
Task comparison		Mean difference	Significance
Uncued	Standard	32.30	0.30
	Reverse	28.19	0.49
	Outcome	23.56	0.71
	Random	35.17	0.24
	Loss	-15.38	0.93
Standard	Reverse	-4.10	1.00
	Outcome	-8.74	0.99
	Random	2.87	1.00
	Loss	**-47.68**	0.02
Reverse	Outcome	-4.63	1.00
	Random	6.98	1.00
	Loss	-43.57	0.06
Outcome	Random	11.61	0.98
	Loss	-38.94	0.14
Random	Loss	**50.55**	0.02

Table S1: Comparisons of risk scores simulated from independent model subject-level parameter estimates using Tukey's HSD test. Bolded values indicate a significant difference.

Table S2: Scaled + offset model simulated risk score comparisons
Tukey HSD

Task comparison		Mean difference	Significance
Uncued	Standard	**38.88**	0.014
	Reverse	24.12	0.348
	Outcome	**36.32**	0.044
	Random	5.57	0.997
	Loss	6.19	0.996
Standard	Reverse	-14.76	0.777
	Outcome	-2.55	1.000
	Random	**-33.30**	0.040
	Loss	**-32.68**	0.047
Reverse	Outcome	12.20	0.910
	Random	-18.55	0.601
	Loss	-17.93	0.636
Outcome	Random	-30.75	0.110
	Loss	-30.13	0.124
Random	Loss	-0.62	1.000

Table S2: Comparisons of risk scores simulated from scaled + offset model subject-level parameter estimates using Tukey's HSD test. Bolded values indicate a significant difference.



β	uncued	standard	reverse	outcome	random	loss
uncued		*		*		
standard	*				*	*
reverse				*		
outcome	*		*		*	*
random		*		*		
loss		*		*		

η^+	uncued	standard	reverse	outcome	random	loss
uncued		*		*		
standard	*		*		*	
Reverse		*		*		
outcome	*		*		*	*
random		*		*		
loss				*		

η^-	uncued	standard	reverse	outcome	random	loss
uncued						*
standard						
reverse						*
outcome						
random						
loss						

m	uncued	standard	reverse	outcome	random	loss
uncued						
standard						
reverse						
outcome						
random						
loss						

Fig. S1: Group-level posterior estimates of basic model parameters. Asterisks within the inset tables mark parameters for which the 95% HDI of the sample difference did not contain zero, indicating a credible difference. For each distribution, the box demarcates the interquartile interval and the whiskers demarcate the 95% HDI.

References

Ahn, W., Haines, N., & Zhang, L. (2017). Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry, 1*, 24-57. doi:10.1162/CPSY_a_00002

Alter, A. (2017). *Irresistible: The rise of addictive technology and the business of keeping us hooked.* New York: Penguin Press.

Amodeo, L. R., McMurray, M. S., & Roitman, J. D. (2016). Orbitofrontal cortex reflects changes in response–outcome contingencies during probabilistic reversal learning. *Neuroscience, 345*, 27-37. doi:10.1016/j.neuroscience.2016.03.034

Barrus, M. M., & Winstanley, C. A. (2016). Dopamine D3 receptors modulate the ability of win-paired cues to increase risky choice in a rat gambling task. *The Journal of Neuroscience, 36*, 785-794. doi:10.1523/JNEUROSCI.2225-15.2016

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*, 7-15. doi:10.1016/0010-0277(94)90018-3

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*, 1-32. doi:10.18637/jss.v076.i01

Cherkasova, M. V., Clark, L., Barton, J. J. S., Schulzer, M., Shafiee, M., Kingstone, A., . . . Winstanley, C. A. (2018). Win-concurrent sensory cues can promote riskier choice. *The Journal of Neuroscience, 38*, 10362-10370. doi:10.1523/JNEUROSCI.1171-18.2018

Chernoff, C. S., Hynes, T. J., & Winstanley, C. A. (2021). Noradrenergic contributions to cue-driven risk-taking and impulsivity. *Psychopharmacology, 238*, 1765-1779. doi:10.1007/s00213-021-05806-x

Dixon, M. J., Collins, K., Harrigan, K. A., Graydon, C., & Fugelsang, J. A. (2013). Using sound to unmask losses disguised as wins in multiline slot machines. *Journal of Gambling Studies, 31*, 183-196. doi:10.1007/s10899-013-9411-8

Dixon, M. J., Harrigan, K. A., Sandhu, R., Collins, K., & Fugelsang, J. A. (2010). Losses disguised as wins in modern multi-line video slot machines. *Addiction, 105*, 1819-1824. doi:10.1111/j.1360-0443.2010.03050.x

Dixon, M. J., Harrigan, K. A., Santesso, D. L., Graydon, C., Fugelsang, J. A., & Collins, K. (2013). The impact of sound in modern multiline video slot machine play. *Journal of Gambling Studies, 30*, 913-929. doi:10.1007/s10899-013-9391-8

Economidou, D., Theobald, D. E. H., Robbins, R. W., Everitt, B. J., & Dalley, J. W. (2012). Norepinephrine and dopamine modulate impulsivity on the five-choice serial reaction time task through opponent actions in the shell and core sub-regions of the nucleus accumbens. *Neuropsychopharmacology, 37*, 2057-2066. doi:10.1038/npp.2012.53

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis* . Boca Raton: Chapman and Hall/CRC.

Goudriaan, A. E., Oosterlaan, J., de Beurs, E., & van den Brink, W. (2005). Decision making in pathological gambling: A comparison between pathological gamblers, alcohol dependents, persons with tourette syndrome and normal controls. *Brain Research. Cognitive Brain Research, 23*, 137-151. doi:10.1016/j.cogbrainres.2005.01.017

Griffiths, M. D. (1993). Fruit machine gambling: The importance of structural characteristics. *Journal of Gambling Studies, 9*, 101-120. doi:10.1007/BF01014863

Haar, C. V. (2020). Challenges and opportunities in animal models of gambling-like behavior. *Current Opinion in Behavioral Sciences, 31*, 42-47. doi:10.1016/j.cobeha.2019.10.013

Hathaway, B. A., Schumacher, J. D., Hrelja, K. M., & Winstanley, C. A. (2021). Serotonin 2C antagonism in the lateral orbitofrontal cortex ameliorates cue-enhanced risk preference and restores sensitivity to reinforcer devaluation in male rats. *eNeuro, 8*, ENEURO.0341-21.2021. doi:10.1523/ENEURO.0341-21.2021

Izquierdo, A. (2017). Functional heterogeneity within rat orbitofrontal cortex in reward learning and decision making. *The Journal of Neuroscience, 37*, 10529-10540. doi:10.1523/JNEUROSCI.1678-17.2017

Jakob, A. M. V., Mikhael, J. G., Hamilos, A. E., Assad, J. A., & Gershman, S. J. (2021). Dopamine mediates the bidirectional update of interval timing. *bioRxiv,* 2021.11.02.466803. doi:10.1101/2021.11.02.466803

Jensen, C., Dixon, M. J., Harrigan, K. A., Sheepy, E., Fugelsang, J. A., & Jarick, M. (2013). Misinterpreting 'winning' in multiline slot machine games. *International Gambling Studies, 13*, 112-126. doi:10.1080/14459795.2012.717635

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience, 21*, 159-RC159. doi:10.1523/JNEUROSCI.21-16-j0002.2001

Kruschke, J. (2014). *Doing bayesian data analysis*. Saint Louis: Elsevier Science & Technology.

Langdon, A. J., Hathaway, B. A., Zorowitz, S., Harris, C. B. W., & Winstanley, C. A. (2019). Relative insensitivity to time-out punishments induced by win-paired cues in a rat gambling task. *Psychopharmacology, 236*, 2543-2556. doi:10.1007/s00213-019-05308-x

Limbrick-Oldfield, E. H., Mick, I., Cocks, R. E., McGonigle, J., Sharman, S. P., Goldstone, A. P., . . . Clark, L. (2017). Neural substrates of cue reactivity and craving in gambling disorder. *Translational Psychiatry, 7*, e992. doi:10.1038/tp.2016.256

Loba, P., Stewart, S. H., Klein, R. M., & Blackburn, J. R. (2001). Manipulations of the features of standard video lottery terminal (VLT) games: Effects in pathological and non-pathological gamblers. *Journal of Gambling Studies, 17*, 297-320. doi:10.1023/A:1013639729908

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 Behavior Processes, 43*, 262-279. doi:10.1037/xan0000141

Murch, W. S., Chu, S. W. M., & Clark, L. (2017). Measuring the slot machine zone with attentional dual tasks and respiratory sinus arrhythmia. *Psychology of Addictive Behaviors, 31*, 375-384. doi:10.1037/adb0000251

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience, 22*, 1544-1553. doi:10.1038/s41593-019-0470-8

Pattij, T., Janssen, M. C. W., Vanderschuren, Louk J. M. J, Schoffelmeer, A. N. M., & Van Gaalen, M. M. (2007). Involvement of dopamine D1 and D2 receptors in the nucleus accumbens core and shell in inhibitory response control. *Psychopharmacology, 191*, 587-598. doi:10.1007/s00213-006-0533-x

Pitchers, K. K., Sarter, M., & Robinson, T. E. (2018). The hot 'n' cold of cue-induced drug relapse. *Learning & Memory (Cold Spring Harbor, N.Y.), 25*, 474-480. doi:10.1101/lm.046995.117

Spetch, M. L., Madan, C. R., Liu, Y. S., & Ludvig, E. A. (2020). Effects of winning cues and relative payout on choice between simulated slot machines. *Addiction (Abingdon, England), 115*, 1719-1727. doi:10.1111/add.15010

Vandaele, Y., Pribut, H. J., & Janak, P. H. (2017). Lever insertion as a salient stimulus promoting insensitivity to outcome devaluation. *Frontiers in Integrative Neuroscience, 11*, 23. doi:10.3389/fnint.2017.00023

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research, 11*, 3571-3594.

Zeeb, F. D., Robbins, T. W., & Winstanley, C. A. (2009). Serotonergic and dopaminergic modulation of gambling behavior as assessed using a novel rat gambling task. *Neuropsychopharmacology (New York, N.Y.), 34*, 2329-2343. doi:10.1038/npp.2009.62

Zeeb, F. D., & Winstanley, C. A. (2011). Lesions of the basolateral amygdala and orbitofrontal cortex differentially affect acquisition and performance of a rodent gambling task. *The Journal of Neuroscience, 31*, 2197-2204. doi:10.1523/JNEUROSCI.5597-10.2011

Zeeb, F. D., & Winstanley, C. A. (2013). Functional disconnection of the orbitofrontal cortex and basolateral amygdala impairs acquisition of a rat gambling task and disrupts animals' ability to alter decision-making behavior after reinforcer devaluation. *The Journal of Neuroscience, 33*, 6434-6443. doi:10.1523/JNEUROSCI.3971-12.2013