Neuron

Article

Activity of Prefrontal Neurons Predict Future Choices during Gambling

Highlights

- Activity of prelimbic neurons predict upcoming decisions during gambling
- Future choice –predictive firing occurs during evaluation of current outcome
- Time-specific inactivation of prelimbic cortex increases high-risk gambling behavior
- Prelimbic neurons contribute to adjusting decisions based on internal valuations

Authors

Johannes Passecker, Nace Mikus, Hugo Malagon-Vina, ..., Gordon Fishell, Georg Dorffner, Thomas Klausberger

Correspondence

johannes.passecker@meduniwien.ac.at (J.P.), thomas.klausberger@meduniwien.ac.at (T.K.)

In Brief

Passecker et al. show that specialized neurons in the prelimbic cortex of rats predict the next choice during the outcome evaluation in a gambling task, even for unlikely or uncertain decisions. Disrupting the prelimbic cortex led to excessive risk taking.

Passecker et al., 2019, Neuron *101*, 152–164
January 2, 2019 © 2018 Elsevier Inc.
https://doi.org/10.1016/j.neuron.2018.10.050

Neuron
Article

Activity of Prefrontal Neurons Predict Future Choices during Gambling

Johannes Passecker,[1,5,*] Nace Mikus,[1,2] Hugo Malagon-Vina,[1] Philip Anner,[1,3] Jordane Dimidschstein,[4,6]
Gordon Fishell,[4,6] Georg Dorffner,[3] and Thomas Klausberger[1,7,*]
[1]Center for Brain Research, Division of Cognitive Neurobiology, Medical University Vienna, Vienna, Austria
[2]Department of Basic Psychological Research and Research Methods, University of Vienna, Vienna, Austria
[3]Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria
[4]NYU Neuroscience Institute, NYU School of Medicine, New York City, NY, USA
[5]Present address: Zuckerman Mind Brain Behavior Institute, Departments of Physiology and Neuroscience, Columbia University, New York City, NY, USA
[6]Present address: Department of Neurobiology Harvard Medical School and the Stanley Center at the Broad
[7]Lead Contact
*Correspondence: johannes.passecker@meduniwien.ac.at (J.P.), thomas.klausberger@meduniwien.ac.at (T.K.)
https://doi.org/10.1016/j.neuron.2018.10.050

SUMMARY

Neuronal signals in the prefrontal cortex have been reported to predict upcoming decisions. Such activity patterns are often coupled to perceptual cues indicating correct choices or values of different options. How does the prefrontal cortex signal future decisions when no cues are present but when decisions are made based on internal valuations of past experiences with stochastic outcomes? We trained rats to perform a two-arm bandit-task, successfully adjusting choices between certain-small or possible-big rewards with changing long-term advantages. We discovered specialized prefrontal neurons, whose firing during the encounter of no-reward predicted the subsequent choice of animals, even for unlikely or uncertain decisions and several seconds before choice execution. Optogenetic silencing of the prelimbic cortex exclusively timed to encounters of no reward, provoked animals to excessive gambling for large rewards. Firing of prefrontal neurons during outcome evaluation signals subsequent choices during gambling and is essential for dynamically adjusting decisions based on internal valuations.

INTRODUCTION

Discoveries of neuronal firing patterns reflecting economic and subjective value (Padoa-Schioppa and Assad, 2006; Kable and Glimcher, 2007), risk-taking (Ogawa et al., 2013), or reward prediction (Schultz et al., 1997) in distinct synaptic circuits (Friedman et al., 2015) have provided mechanisms and inspired several models for value-based decision making (Rangel et al., 2008; Glimcher and Fehr, 2014; Padoa-Schioppa, 2011; Sugrue et al., 2005; Hunt and Hayden, 2017). In contrast to a standard behavioral task design, many of our decisions are not guided by external perceptual cues informing us about a correct or an incorrect choice, and decisions are not often based on perceptually presented stimuli with a deterministic consequence. Regularly, choices need to rely on experience-based inner valuations of different options with a probabilistic outcome distribution. Gambling tasks with changing reward contingencies serve as a model in which flexible decision making relies on internal valuations without external cue guidance and aims toward reward maximization and individual satisfaction. During gambling, encounters of choice options under uncertainty often lead to seemingly unpredictable decisions, and the neuronal mechanisms driving such unguided decision-making based on the inner valuation of probabilistic outcome remain poorly understood.

Neuronal signatures of economic choice have been reported in the lateral orbitofrontal cortex (Padoa-Schioppa and Assad, 2006; Padoa-Schioppa, 2011). In rodents, recent evidence has emerged that medial parts of the prefrontal cortex may be paramount (Gardner et al., 2017), but distinct neuronal underpinnings are yet to emerge. A series of findings has identified the prelimbic cortex as a key structure in value-guided decision making (Zeeb et al., 2015; St Onge and Floresco, 2010; Balleine and Dickinson, 1998), which is in line with findings linking the medial prefrontal cortex with the top-down cognitive control based on internal valuations across species (Koechlin et al., 2003). The prelimbic cortex has also been suggested in contribution to behavioral flexibility, which enables adaptive control and allows spontaneous choices based on internal valuation (Dolan and Dayan, 2013; Kolling et al., 2014). During choices under risk, dopaminergic cells have been reported (Stauffer et al., 2014) to reflect the utility function of decisions. We aimed to unravel neuronal signals in the prelimbic cortex of rats that combine various relevant signals and reflect a binary choice output during gambling. For this purpose, we adopted a two-arm bandit-task design to incite inherent valuation processes for decision optimization during dynamically changing gambling conditions. We measured and manipulated neuronal activity in the prelimbic cortex of rats to uncover firing patterns, which, in the absence of perceptual cues or offers, signal upcoming decisions based on internal valuation of stochastic past experiences.

RESULTS

Rats Can Successfully Adjust Their Decisions to Maximize Long-Term Amount of Reward during a Gambling Task

Inspired by bandit tasks for humans, we trained rats to choose freely and without cue guidance between a certain-small reward on the ''safe-arm'' of a Y-maze or a possible-big reward on the ''gamble-arm'' (Figure 1A). The likelihood of reward on the gamble-arm was changed twice during a session, altering the advantage between the two arms or allocating similar merit to both options (Figures 1B and S1A). We observed that animals were able to adjust their choices to maximize the long-term amount of reward and follow the changes in reward contingencies (Figures 1C, 1D, and S1B–S1E). Choice behavior of animals adapted based on the diverse reward experiences during individual behavioral sessions (Figures S1B and S1C). Comparing the animals' behavior to an optimal agent allowed only a measure of performance but did not allow adequate tracking of subjective values and goal preferences during individual behavioral sessions (Table S1). Among several tested behavioral models (Table S1), we applied a reinforcement-learning (RL) model (Sul et al., 2010) to estimate subjective goal values and predicted 80.4% (±2.7 SEM) of the animals' choices. We refer to the modeled probability for a subsequent choice of the gamble-arm as ''choice evidence for gamble'' (Figure 1B). Modeled choice evidence allowed a more refined representation of subjective goal value changes during the task (Figures S1B and S1C). As expected, we observed fewer choices for the gamble-arm during episodes of low choice evidence for gamble (Figure 1E). During periods with ambiguous choice evidence, rats adjusted their strategy more often and made significantly more changes between the two arms (Figure 1F), while running speed did not correlate with different levels of choice evidence (Figures S1F and S1G).

Firing Patterns of Prelimbic Neurons and No-Reward Activated Cells During Performance of the Gambling Task

We performed tetrode recordings in four rats during task performance and measured the activity of 1,006 neurons across 45 behavioral gambling sessions. A demixed principal-component analysis (DPCA) (Kobak et al., 2016) and a multiple regression analysis (Figure 2) revealed that neuronal firing in the prelimbic cortex differentiates according to different task episodes, occurrence or absence of reward, and according to modeled choice evidence. Strikingly, a major proportion of recorded neurons in the prelimbic cortex significantly increased their firing during the experience of no-reward at the gamble-arm. The firing of these classified no-reward activated cells (n = 402) was significantly correlated to the occurrence of no-reward during any three consecutive time-bins during the reward episode (Figure 2B, right panel). Additionally, these cells changed their firing according to choice evidence and exhibited higher firing rates during trials with low- and ambiguous- compared to high-choice evidence for gamble (Figures 3A–3D and S2A–S2C). This distinction in firing rate according to different levels of choice evidence was restricted to the gamble-arm and was not observed during safe-arm choices, which were always rewarded (Figure 3B). The increased firing rate during low and ambiguous choice evidence for gamble, could not be explained by possible speed-related changes and thus unlikely reflect possible motivational biases (Figures S1F and S1G). No-reward-activated neurons were similarly present across animals (~40%, Figure S2D); their firing reflected current but not past reward information (Figure S2E) and allowed a correct prediction about reward occurrences on 90.97% ± 1.60% (mean% ± SEM) of trials (Figure S2F). Although the firing of some no-reward-activated neurons was not modulated by choice evidence (Figures 3D and S2D), the activity of most of these cells exhibited additional correlation with choice evidence and arm/or chosen goal in a task-episode-dependent manner (Figures 3C, 3D, and S2A–S2C), likely integrating context, goal value, and reward as reported earlier (Mante et al., 2013; Rigotti et al., 2013; Raposo et al., 2014).

No-Reward Activated Neurons Provide Limited Accuracy in Future Choice Prediction

Observing an influence of primary task and decision variables on the prelimbic neuronal activity during the gambling behavior, we asked whether prelimbic firing patterns might be predictive for future choices during task performance. We applied an elastic-net regression as feature selector to identify the best predictors and evaluated their power with a general linear prediction model. Using firing rates of either all recorded prelimbic neurons or the activity of no-reward activated cells only as input for the regression, the resulting models, on average, correctly predicted 78.1% ± 1.3% and 75% ± 1.3% (mean ± SEM), respectively, of future choices across all behavioral conditions (Figures S3A–S3C). In order to explore which neuronal signals might be responsible for successful predictions, we focused on a key situation during gambling, when an animal chooses the gamble-arm but does not receive a reward. What will the animal decide to do in the next trial: continue gambling or play it safe? In this situation, the animals choose the safe arm in the subsequent trial with almost similar likelihood 45% ± 5% (mean ± SEM) as the gamble-arm. This intriguing scenario allowed us to control for goal-arm location and reward information as in all instances the animal is located on the same goal arm and receives no reward, while retaining high unpredictability about what choice the animal will cast in the following trial. We observed that changes in track trajectory, head direction (relative speed), and heading (degrees), during the reward episode were independent of the choice the animal will cast on the subsequent trial (see Figures S4A–S4C). We analyzed whether the firing of no-reward-activated cells in such a scenario of non-rewarded gamble-arm trials is indicative of future choice. During periods of high-choice evidence for gamble, these cells fired with higher rates when the animal will change its strategy and select the safe-arm in the subsequent trial, compared to no-reward encounters, when the animal will decide to continue gambling on the next trial (Figure 4A). Indeed, no-reward activated cells predict future choices under such conditions and provided more predictive information toward a change in strategy than the other recorded cells (Figures 4B–4D). The accuracy of the prediction increased with increasing number of co-recorded cells (r = 0.374, p = 0.023). However, during periods of ambiguous choice evidence for

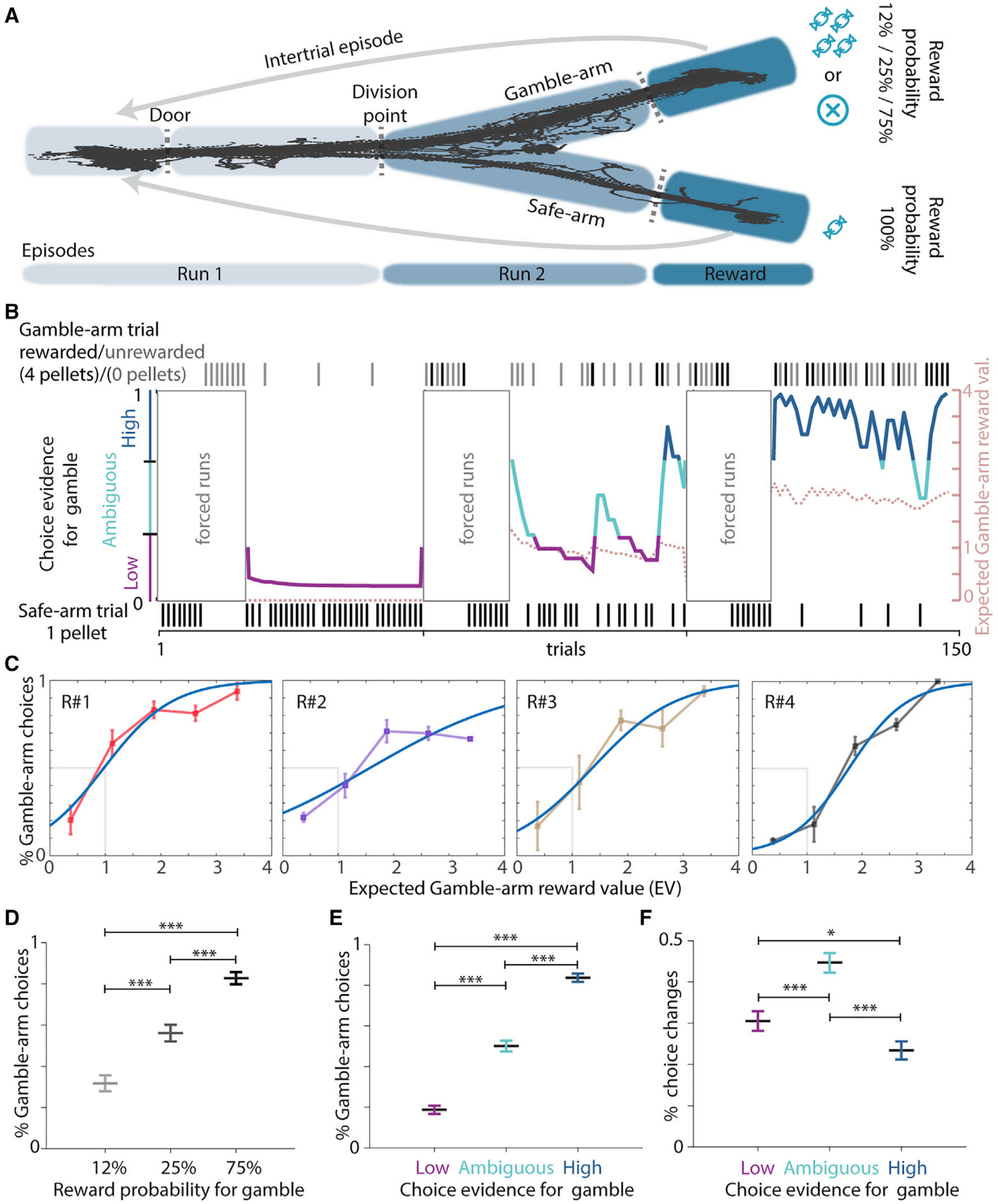

Figure 1. Behavioral Analysis for the Gambling Task

(A) Running on a Y-maze, rats choose between a small and always-available or a big reward given only with a 12%, 25%, or 75% probability.

(B) For an individual session, choice evidence for going to the gamble-arm was calculated with a reinforcement learning model and classified as low (purple), ambiguous (turquoise), or high (blue). Animals had to explore both arms in forced trials at the beginning of each block of trials with a defined reward probability on the gamble-arm. Dotted line: expected value for the gamble arm based on reward occurrence. Ticks indicate observed choice for each trial.

(C) Each animal (R#1–R#4) was able to maximize reward (see also Figure S1A) according to reward occurrence on the gamble arm (blue curve: logistic fit function; gray: only Gamble-arm choices above an expected gamble arm reward value [EV] of 1 are optimal).

(D) Increasing reward probability leads to increased preference for the gamble-arm (one-way ANOVA $F_2 = 50.529$, $p < 0.001$; $n_{12,5} = 45$, $n_{25} = 43$, $n_{75} = 45$ sessions).

(E) Modeled choice evidence reflects behavior of animals (one-way ANOVA: $F_2 = 178.703$, $p < 0.001$).

(F) Changes in arm choice were most abundant during ambiguous choice evidence (one-way ANOVA: $F_2 = 23.511$, $p < 0.001$).

Data as mean ± SEM, post hoc multiple comparison Student-Newman-Keuls method $^*p < 0.05$; $^{***}p < 0.001$, $n_{amb} = 45$, $n_{low} = 39$, $n_{high} = 45$ sessions for (D) and (E), 4 rats for (C), (E), and (F).

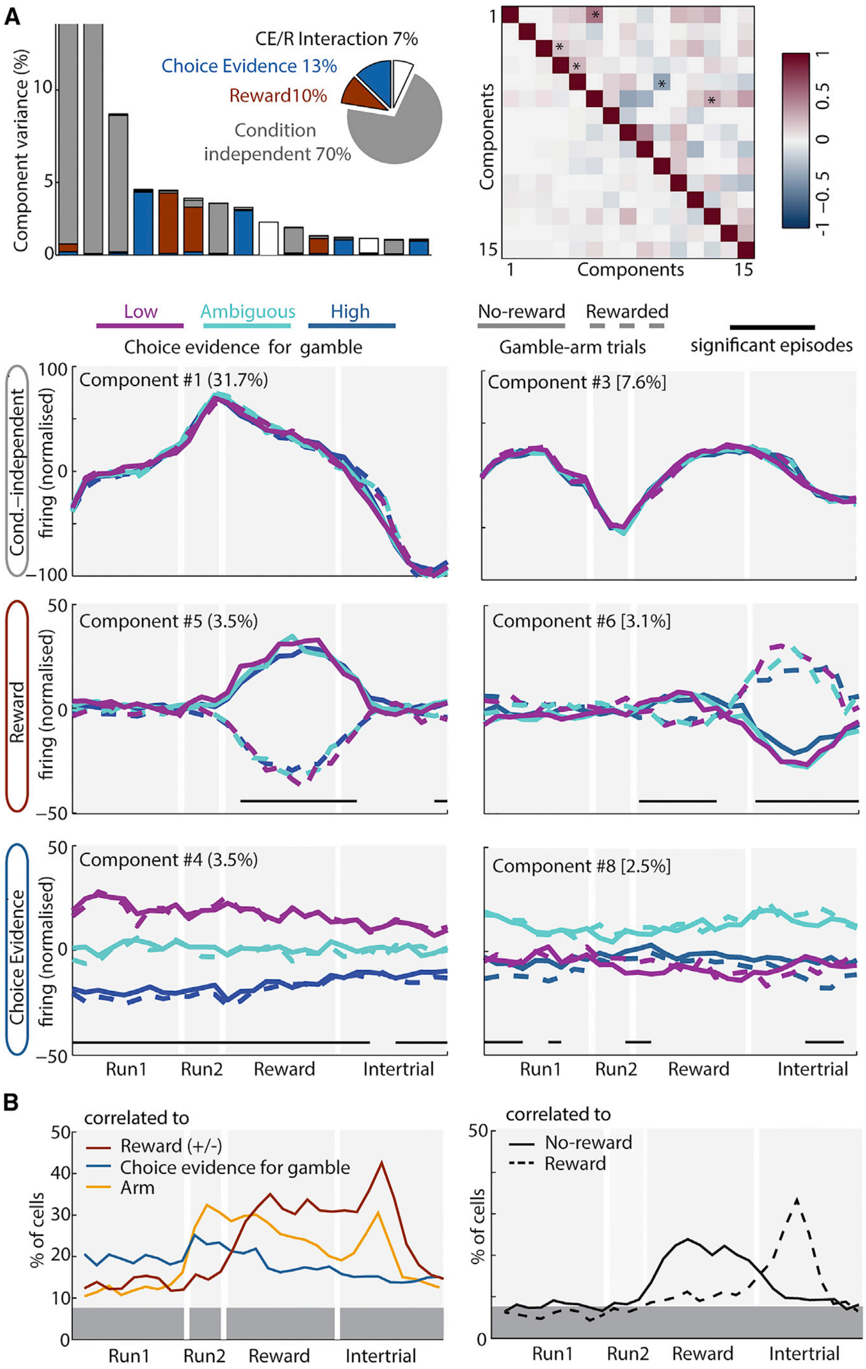

Figure 2. Modulation of Prelimbic Neuronal Activity by Reward, Choice Evidence, and Task Episode

(A) Demixed principal-component analysis (PCA) reveals major contributors of firing rate variance on gamble-arm trials. Time and episode modulation contributes most significantly to the firing rate variances followed by choice evidence and reward as denoted in the pie chart segments. The first 15 principal components of the demixed PCA and its contributing variables for gamble-arm firing rate modulation are depicted in the bar graph (left top panel). The first two component contributions are cut for comparison reasons. The upper-right triangle of the top right panel depicts dot products between all pairs of the first 15 demixed principal axes. Stars denote significantly non-orthogonal principal components; note components 5 and 4 indicating an interaction of choice evidence- and reward omission-related neuronal activity. Bottom left triangle shows correlations between all pairs of the first 15 principal components. A selection of the main principal components indicates time-dependent reward modulation (C. #5 and #6), choice evidence modulation (C. #4 and #8), and task episode modulation (C. #1 and #3). Black lines indicate significant periods.

(B) Multiple regression analysis indicates correlations of reward occurrence, modeled choice evidence, and spatial arm location with firing of prelimbic neurons along trial episodes (left panel). The firing of large subsets of prelimbic neurons is correlated with reward omission or reward occurrence in a time-dependent manner (right panel).

Note: to be classified as a no-reward activated neuron, firing rate had to depict significant correlation in three (out of 9) consecutive time bins during the reward episode. Dark gray depicts chance level.

gamble (with multiple choice fluctuations), no-reward activated cells, on average, did not significantly differentiate their firing rate according to future choices in upcoming trials (Figure 4A, right panel). Thus, the predictive power of no-reward activated cells appears at least partly linked to the previously described correlation of firing rate with choice evidence. Therefore, the firing of no-reward activated cells, as a population, may not allow a reliable prediction of subsequent trial choices on a trial-by-trial basis or during periods with ambiguous choice evidence.

A Differentiating Firing of Choice Predicting Cells during the Encounter of No-Reward Indicates the Choice of the Animal in the Subsequent Trial

To adjust for choice evidence-modulated firing of no-reward activated cells, we subtracted the mean firing rate of no-reward activated cells during trial t−1 (excluding the reward episode) from the firing rate during reward episode of trial t. Using these relative firing rate values as input predictor for future choice in trial t+1, the elastic-net regression selected a population of cells, whose actually-recorded firing patterns exhibited stable high predictive power for future choices. Timed to the encounter of no-reward on the gamble-arm, these cells significantly differentiate their recorded firing rate according to the subsequent choice of the animal (Figures 5A–5D, S5A, and S5B). Even during periods of ambiguous choice evidence or for unlikely upcoming choices (safe-arm choices during periods

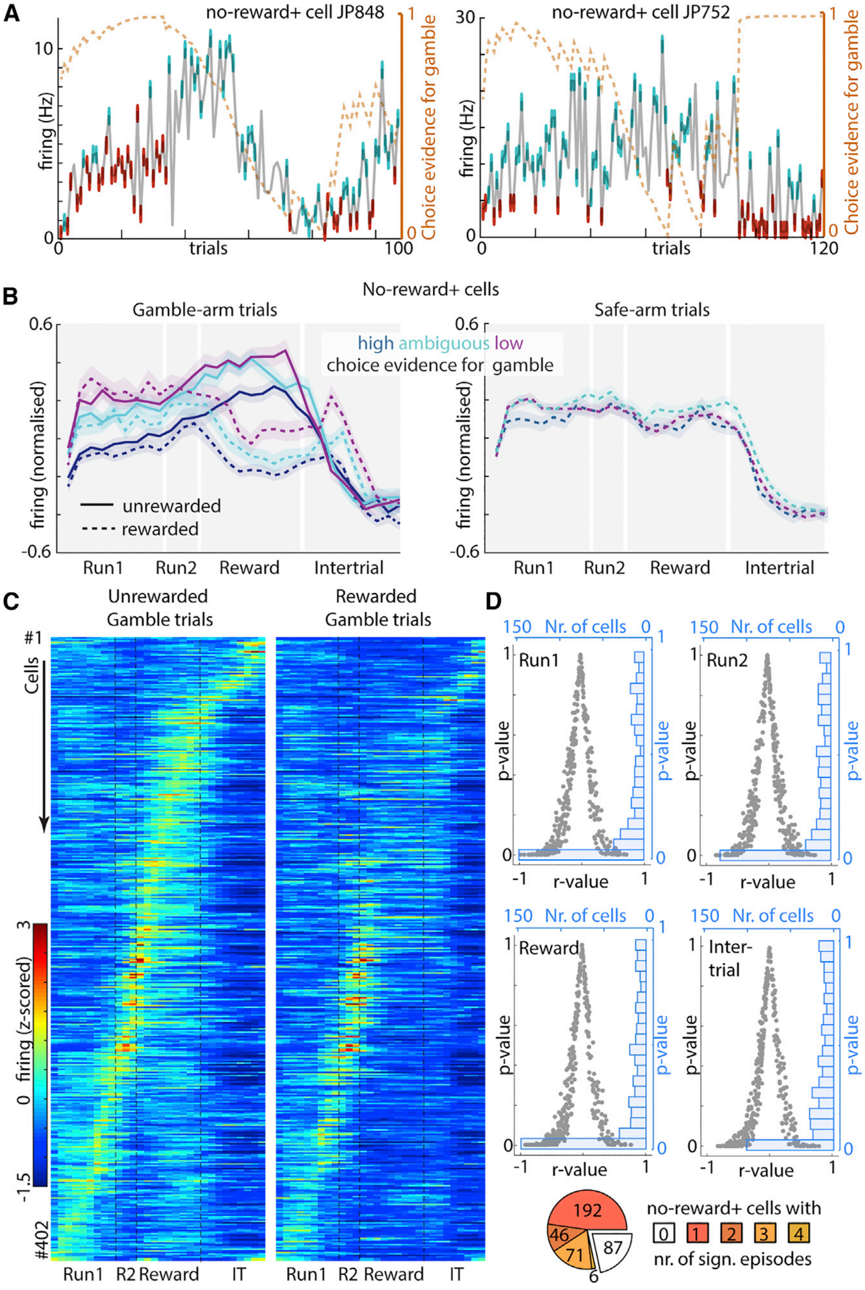

Figure 3. Firing of No-reward Activated Cells Reflects Reward Occurrence and Choice Evidence

(A) Firing rate during the reward episode of two no-reward activated neurons increases during unrewarded gamble trials (blue) and depends on choice evidence. Red: rewarded gamble-arm trials.

(B) Firing of no-reward activated cells (n = 402) according to reward occurrence, modeled choice evidence, and arm choices. Note an increased firing during non-rewarded trials and during periods of ambiguous and low choice evidence for gambles exclusively on the gamble-arm.

(C) Normalized firing of no-reward activated neurons for unrewarded (left, sorted for peak firing) and rewarded (right) gamble-arm trials. Note, cells with peak firing during but also outside of the reward episode differentiate firing during the reward episode according to reward occurrence (see Figure S2B for individual examples; for visualization purposes, maxima and minima outside the color range are omitted).

(D) Correlations between firing rate and choice evidence during four indicated task episodes (unrewarded trials only, 402 neurons). Bottom: large fractions of no-reward activated cells exhibit significantly correlated firing (unrewarded trials) with choice evidence for gamble arm for at least one trial episode (run1, run2, reward, and/or intertrial episode, n = 402).

of high choice evidence for gambling), these “choice-predicting cells” exhibited significantly higher firing during the reward episode when the animal would change its strategy in the next trial and select the safe option as opposed to choosing a further gamble subsequently (Figures 5A, middle panel, 5B, 5C, and S5B). We observed that this firing differentiation was independent from reward, location, motoric confounds, different levels of value, choice evidence, and reward prediction error (Figures S4A–S4C and S5F–S5H). Nevertheless, many of the latter variables present trial-to-trial variations, which could still account for firing rate variances in the investigated trial scenarios. Thus, we used the residuals of a multivariate regression (controlling against reward prediction error [RPE], action value for choosing the gamble-arm, trajectory changes, head-directional changes and choice evidence) instead of the firing rate as input to an elastic net regression analysis. The firing of the selected cells differentiated their firing according to future choice (Figures 5D and S5K), confirming that future choice prediction of these cells is independent of reward prediction error, action value of the gamble-arm, trajectory changes, head-directional changes, and choice evidence. Furthermore, the identification of choice-predictive signals remained independent of different reinforcement models used (alternative RL model with similar predictive power; Figure S5I) and independent of reinforcement model parameters during ambiguous choice situations when expected reward values were similar on both arms (Figure S5F) or probability of gamble is close to 0.5 (Figure S5G).

If neurons indeed present a firing rate differentiation during reward evaluation dependent of future choice but independent of changes in goal value and reward prediction error, then they should continue to do so independent of prior experiences. First, we confirmed that the outcome during the trial before the non-rewarded gamble trial has no significant impact on the firing rate differentiation of choice-predictive cells (Figure S5C). Then, we

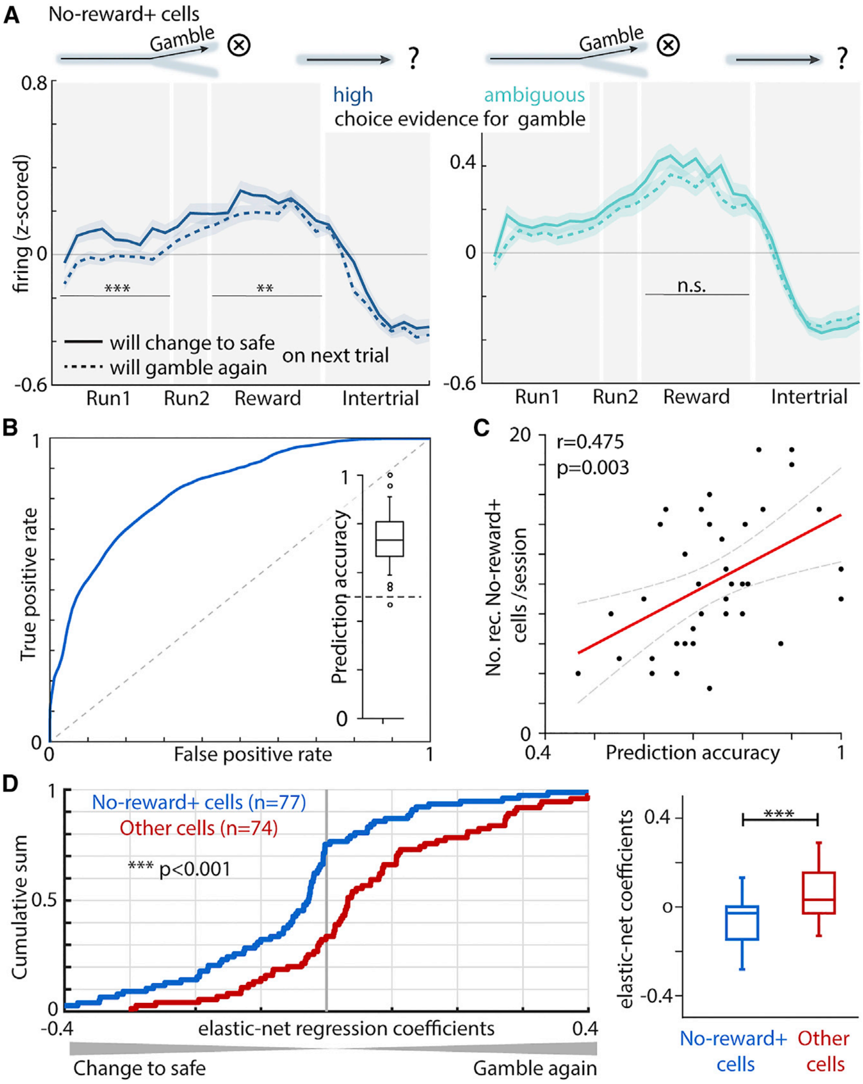

Figure 4. Predictions of Future Choices Based on the Firing of No-Reward-Activated Neurons

(A) During non-rewarded gamble-arm trials, no-reward activated cells differentiate their average firing according to the subsequent choices of the animal for trials with high but not ambiguous choice evidence for gamble. Note: the difference in absolute firing rate between left and right panels contributes to prediction power. Wilcoxon signed rank test, alpha = 0.00167 (bonferroni corrected); left: reward episode: Z = −2.864, p = 0.0261; all other episodes n.s; n = 308; right: run1 episode: Z = −5.533, p < 0.0001; Reward episode: Z = −3.245, p = 0.0012; all other episodes n.s; n = 339.

(B) Receiver-operating statistics of successful predictions of future choices during non-rewarded gamble trials based on the firing of no-reward activated cells (mean curve ± SEM). Inlet: prediction accuracy per session mean = 74.2%; median = 73.3%.

(C) Prediction accuracy of the model increases with higher numbers of simultaneously recorded no-reward-activated cells.

(D) No-reward-activated cells (blue, n_{nor} = 77) better predict the change of choice compared to other recorded cells (red, n_{other} = 74) as indicated by their elastic-net coefficients after non-rewarded gamble trials. Left panel: two sample Kolmogorov-Smirnov test; right panel: Mann-Whitney U test: U = 1,676.000, p < 0.001; (A–C): n = 37 sessions).

analyzed recurring choice scenarios during four consecutive trials, with defined outcomes during the first three trials, and the firing rate of choice-predicting cells was analyzed dependent on what choice the animal will cast on the fourth trial (Figures 5E, S5D, and S5E). Irrespective of whether the animal has experienced none or repeated reward omissions in the two previous trials, the choice-predicting cells always differentiated their firing in the third trial according to animal's choice in the fourth trial. This confirms that previous outcomes and choices have no influence on the choice-predictive signal, suggesting its independence from expectancy and surprise. Within this population of choice-predicting cells, the predictive increase in firing is temporally restricted to the reward episode and does not persist significantly earlier or later in the non-rewarded trial (Figures S5D and S5E).

In contrast to the reinforcement learning model, which is based on behavioral parameters only, a prediction model based on the firing of these choice-predicting cells achieved accurate persistent forecasts of future choices even during ambiguous choice evidence (Figure 5F). In fact, the predictive power of the no-reward activated cell population benefited from the population of choice-predicting cells (Figure S5L).

Next, we tested whether the predictive neuronal activity of those choice-predicting cells might also allow inference about future choice when the animal consumes a small reward on the safe arm. However, these cells did not exhibit a differentiating firing pattern for future choice during safe arm trials (Figure 5G). Using an elastic-net regression, we identified a distinct population of 88 cells, which significantly differentiated their firing rate during the reward episode of safe-arm trials depending on the choice of the animal in the following trial during ambiguous- and low-choice-evidence for gambling (Figure 6A). These neurons did not carry predictive power for upcoming choices during non-rewarded gamble-arm trials (Figure 6B). Thus, distinct sets of neurons in the prelimbic cortex provide a reliable and predictive firing-rate-based signal, indicating the upcoming choice on the following trial in an arm-specific manner. This aligns with the observation that neurons in our task design present strong arm-identity-dependent information (Figures S6A and S6B). When gamble and safe arm identity was switched halfway through the task (without physically moving the goal arms), neurons, in general, did not respond to changes in spatial location of the goal arms but maintained their firing according to gamble or

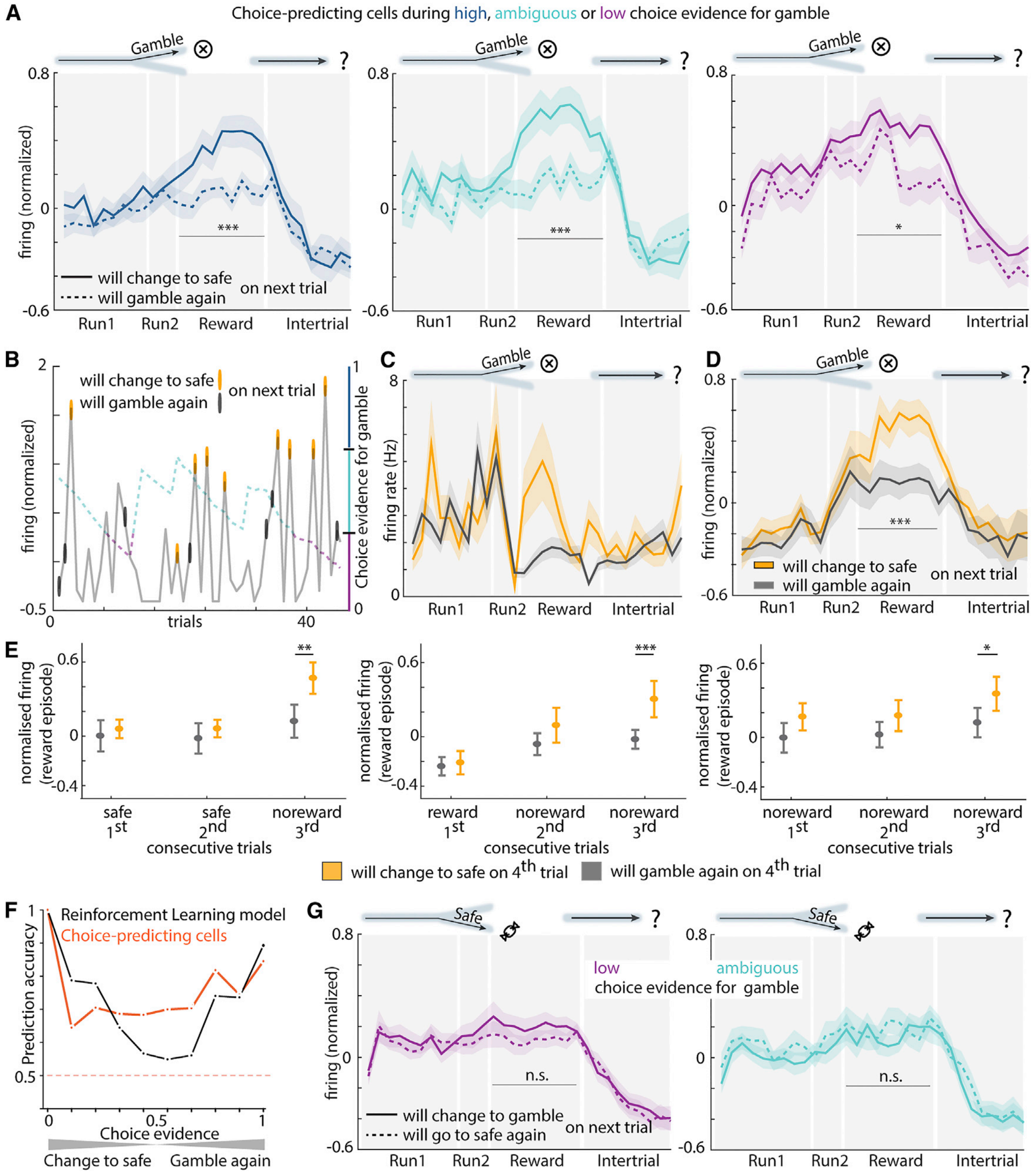

Figure 5. Prediction of Future Decisions by the Firing Rate of Choice-Predicting Cells during Evaluation of Negative Outcomes

(A) At the time of no-reward, firing of choice-predicting cells indicates rat's choice in the next trial even during ambiguous choice evidence, before unlikely choices for the safe arm during high choice evidence for gambling, or for unlikely choices for the gamble arm during low choice evidence for gamble. Neurons significantly increase firing during the occurrence of no-reward on gamble-arm trials before the animal will change its strategy to the safe-arm in the next trial compared to a subsequent gamble-arm choice. (Signed rank test, adj. for multiple comparisons, ***$p < 0.0001$; left panel: $Z = -5.9879$, right panel: $Z = -5.0817$). Note that only unrewarded gamble-arm trials are considered here.

(B) Firing of a choice-predicting cell during reward episodes across 48 consecutive trials with mostly ambiguous choice evidence. Gold and black ticks indicate no-reward occurrence on the gamble-arm with a safe-arm or gamble-arm choice in the next trial, respectively.

(C) Different visualization of a choice-predicting cell across all non-rewarded gamble arm-trials (see Figures S5A and S5B for more examples).

(D) A multivariate regression of the firing rate of choice-predicting cells against variance changes of major task variables (choice-evidence, reward prediction error, head-direction, movement, and action value of the gamble arm) was performed. The resulting residuals were subjected to a lasso regression analysis. The firing rate of the selected cells (shown here) maintains a significant difference ($p = 1.83e^{-4}$; $n = 20$ sessions) for distinct future choices, indicating their independence from these task variables.

(legend continued on next page)

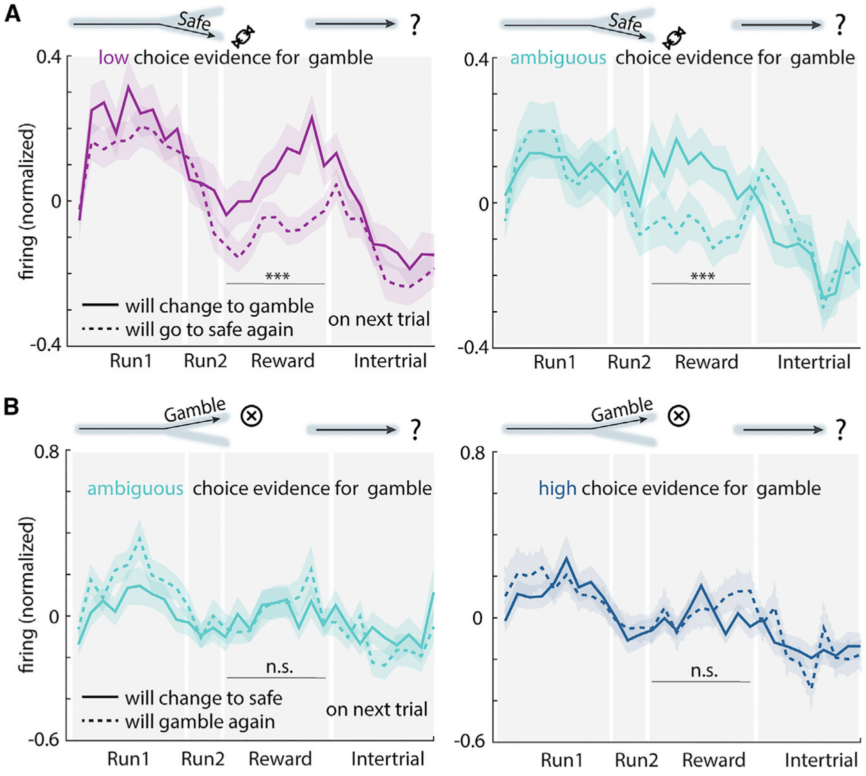

Figure 6. During the Encounter of Reward on the Safe Arm, the Firing of Another Subset of Prelimbic Neurons Indicates the Upcoming Choice of the Animal

(A) Elastic-net regression identified 88 cells which significantly differentiate their firing rate during the reward episode of safe-arm trials depending on the choice of the animal in the following trial. Low-choice evidence: Z = −5.7220, ***p < 0.001, ambiguous choice evidence: Z = −5.0456, ***p < 0.001; n = 81 cells (left) and 86 cells (right) alpha at 0.00166 (Bonferroni corrected).

(B) During unrewarded gamble arm trials, these neurons did not differentiate in their firing for distinct future choices (left panel: choice ep. Z = −2.6146, p = 0.0089, reward ep.: Z = −0.6321, p = 0.527; n.s. n = 81, right panel: reward ep.: Z = −1.2040, p = 0.2286, n = 79).

safe arm identity (Figure S7A–S7F). This further confirms that firing of prelimbic neurons was related more to task and cognitive content rather than to spatial or motoric parameters.

Inactivation of Prelimbic Neurons, Exclusively Timed to the Experience of No-Reward, Increases Gambling Behavior

As prelimbic firing patterns during the occurrence of no-reward are predictive for subsequent choices, we tested whether optogenetic silencing of the prelimbic cortex, exclusively timed to no-reward encounters, impeded optimal decision making. We used a novel viral approach to express channelrhodopsin2 exclusively in GABAergic neurons (Dimidschstein et al., 2016) of the prelimbic cortex (Figures 7A and S8K). Shining blue light via optic fibers into the prelimbic cortex for 1 ms at 66 Hz activated putative interneurons and inhibited the activity of 90% of prelimbic neurons (Figure 7B). Such bilateral and spatially restricted optogenetic silencing (Figures S8A–S8D and S8K) of the prelimbic cortex during task performance and timed only to no-reward encounters during gamble trials impaired the performance of rats (Figures 7C and 7D). Rats persisted in choosing the gamble-arm even during highly unfavorable reward contingencies compared to optimal choice selection with control stimulations during the run1 episode, during the reward episode, of rewarded-gamble or safe-arm trials, or without stimulation (Figures 7E–7H). The significantly increased number of gambles for prelimbic silencing during no-reward resulted in an increased number of disadvantageous decisions as animals continued to gamble for big rewards after non-rewarded trials even in situations when the evidence would indicate the opposite (Figures 7E–7G and S8E). This alteration in choice behavior resulted in fewer arm changes (Figure S8F) and was reflected in parameters of the expected value and reinforcement learning model (Figures S8G and S8H). The increased tendency to take riskier choices might not result from reward-inducing effects of optogenetic interventions as further control experiments show that animals do not prefer locations where such optogenetic stimulations occur (Figures S8I and S8J). Thus, the firing of prelimbic neurons during no-reward encounters is required for adjusting decisions based on negative feedback.

DISCUSSION

Our adapted framework of a two-arm bandit-task required animals to explore and integrate probabilistic reward outcomes and, accordingly, adapt policy selection to strive toward reward maximization. During this task, animals often exhibited volatile choice behavior, and the experience of a negative outcome on the gamble arm presented the animal regularly with an ambiguous decision scenario on whether to choose again the gamble arm or

(E) Predictive firing rate differentiation was not influenced by differences in choice and outcome of the two preceding trials (two-way RM ANOVA, all group and time variables significant; left: p = 0.002, middle: p = 4.07e^{-4} right: p = 0.015; 1st and 2nd trial comparisons all n.s., mean ± standard deviation; see Supplemental Information and Figures S5D and S5E).

(F) A prediction model based on the firing of choice-predicting cells maintains good performance even during ambiguous choice evidence (0.5) (elastic-net input: n = 402; GLM input: n = 151). Dotted line represents chance level.

(G) On the safe arm, the firing of these choice-predicting cells does not indicate the upcoming choice in the next trial (n = 84; same statistics as in A, all episodes n.s.).

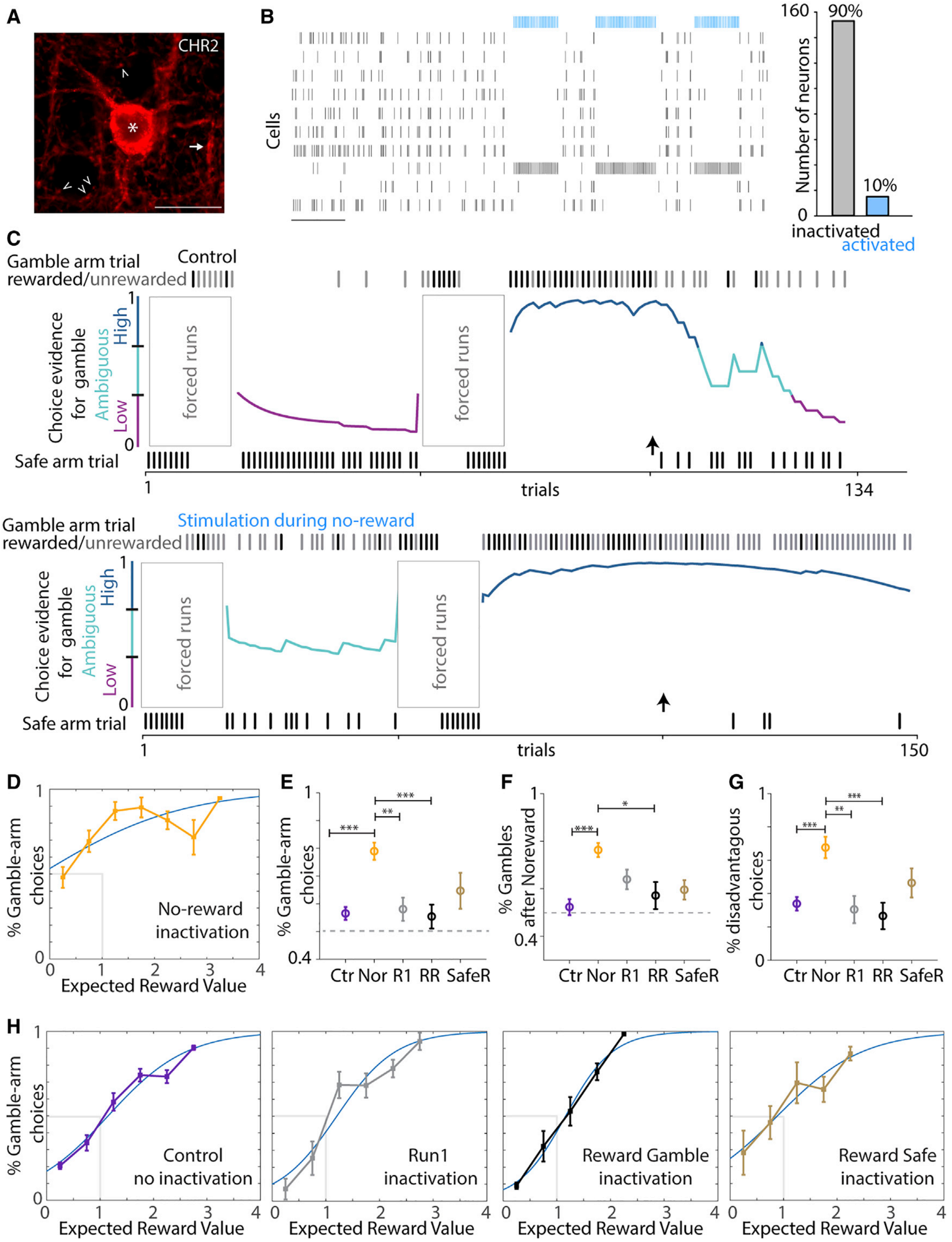

Figure 7. Inactivation of Prelimbic Neurons during No-Reward Increases Disadvantageous Gambling Arm Choices

(A) Viral expression of channelrhodopsin2 in somata, axon terminals (arrowheads), and dendrites (arrows) of prelimbic GABAergic neurons. Scale bar, 20 μm.

(B) Optogenetic activation of a putative GABAergic interneuron (#3 from bottom) and inhibition of putative excitatory principal cells with a 1-ms-long, 66 Hz stimulation protocol.

(C) Compared to control, bilateral optogenetic silencing of the prelimbic cortex, exclusively timed to the occurrence of no-reward during gamble-arm trials, increased gamble-arm choices when safe-arm choices would be favorable (first block) and after an unannounced decrease in reward probability on the gamble arm (arrow).

(legend continued on next page)

the safe arm in the next trial. Taking advantage of this behavioral scenario, we discovered neuronal firing activity in the prelimbic cortex, which informs differentially about future choices of the animal in economic decision making. We discovered a specialized subset of neurons, whose firing patterns during the evaluation of negative outcome signals choice in the subsequent trial even for unlikely choices and more than 5 s before this decision is executed. Optogenetic inhibition of prelimbic activity, exclusively timed to the encounter of no-reward, resulted in an increased number of gambles and confirmed the importance of the underlying prelimbic neuronal activity for optimal decision making. These results suggest reliable reflection of intrinsic evaluation processes of negative outcomes by prelimbic neurons signaling future choice. These signals are well suited for optimizing action adaptations, in particular during conflicting choice situations.

Activity of choice-predictive neurons, whose firing rate increases during reward omissions and indicate future choice, cannot be attributed to classic RPE signals (Schultz et al., 1997) and/or surprise signals (Hayden et al., 2011). During non-rewarded gamble arm trials (outcome is always worse than expected, but negative RPE differentiates in weight), choice-predictive firing patterns remain stable during either low or high levels of negative RPE (see Figure S5H) and during trial-by-trial variations of RPE (see Figure 5D). Different levels of risk (McCoy and Platt, 2005) or the amount of uncertainty in choice outcomes (Kepecs et al., 2008) do not account for the observed choice-predictive firing either (see Figures 5A, 5D, S5F, S5I, and S5J). Independent of estimated subjective (e.g., action values, choice evidence; Tsutsui et al., 2016; Figure 5D) or objective task parameters (e.g., probability of reward; Figure S5G), the choice-predictive firing signal remains across different scenarios of past trial history (Figures 5E and S5C). Major motoric differences in behavior, which are correlated with neuronal firing rates in the medial prefrontal cortex (Lindsay et al., 2018), can also be excluded as confounding factors (Figures 5D, S4, and S5K).

The predictive firing-rate increase of choice-predicting cells precedes an upcoming change to the safe arm on the next trial by several seconds and is highly time restricted to the encounter of no-reward. Thus, they contrast with working-memory signals observed in the prelimbic cortex in tasks where animals are required to hold goal relevant cue information in memory (Fujisawa et al., 2008). There is no evidence that choice-predicting cells in the prelimbic cortex remain informative for subsequent choices for prolonged periods lasting into the next trial (see Figure S5E). Furthermore, the temporal span of at least 5–6 s between firing rate differentiation and decision manifestation and the resulting mix of motoric behaviors during those time windows (transfer, waiting time, and run initiation) makes it highly unlikely that the signal corresponds to preparatory pre-action or pre-motoric signals (Hare et al., 2011; Svoboda and Li, 2018).

Future choice prediction has been reported for a cohort of neurons in the anterior cingulate cortex (Ito et al., 2015). The predictive firing rate of these cingulate neurons peaked just before the execution of the decision of which goal arm to choose. This would be comparable to the run1 episode in our task, suggesting different signals for choice execution in cingulate cortices and choice-predictive signals for the subsequent trial in prelimbic cortices described here during reward evaluation. Decisions in our task design mainly reflect changes in internal goal valuations—they are not associated with manipulations of external perceptual and sensory cues—informing about optimal choice as in cue-dependent and perceptual-decision tasks (Kim and Shadlen, 1999; Matsumoto et al., 2003). Similarly, seminal work from Padoa-Schioppa and colleagues describe neuronal responses in the primate lateral OFC signaling the upcoming “chosen juice” following stimulus presentation (Padoa-Schioppa and Assad, 2006). The cue presented during each trial distinctively informed the animals about goods and outcome. In light of our data, it might be interesting to explore how those neurons in OFC of monkeys fire when subjective good values are presented in an almost equally matched manner and with a stochastic outcome distribution.

Neuronal activity patterns of choice-predicting cells are somewhat reminiscent of cells observed in the posterior cingulate cortex and cingulate motor areas (Hayden et al., 2008; Shima and Tanji, 1998) in primates. These reported neurons increased firing following reward, indicative of shifts in goal choice for upcoming trials. In the case of the posterior cingulate cortex (Hayden et al., 2008), neurons increased their firing independent of the direction of subsequent choice changes, which contrasts with our findings in the prelimbic cortex. Despite differences in task design and species, it will be interesting to understand synaptic interactions and information flow between (and within) the different areas involved in value-guided decision making. Downstream motoric structures are possible candidates to initiate behavioral responses (Kim and Shadlen, 1999; Matsumoto et al., 2003; Kerns et al., 2004; Svoboda and Li, 2018; Barack et al., 2017). The primate cingulate cortex (Pribram et al., 1962) has been linked to error, reward- and value-based decision-making (Rushworth et al., 2011), and conflict management (Mansouri et al., 2009) for optimal choice (Kennerley et al., 2006; Botvinick et al., 2004). The difficult question of species homology for prefrontal cortex highlights the importance of discovering distinct neuronal activity patterns to establish better predictive validity for why failing or shifting reward integration mechanisms result in detrimental decision making across a range of psychiatric conditions (Fujimoto et al., 2017).

The activity of choice-predicting cells suggests that the prelimbic cortex is crucial beyond goal value comparisons for future choice (Sul et al., 2010; McCoy and Platt, 2005) and might contribute to executive signals for optimized decision making.

(D–G) Inactivation of prelimbic cortex, timed to the experience of no-reward (Nor), increased number of gambles (D) compared to controls (E: $F_4 = 8.403$, $p < 0.001$). Ctr, no stimulation; R1, stimulation during run1 episode; RR, stimulation during reward experience on the gamble arm; SafeR, stimulation during reward episode on the safe arm. Compared to controls a persistent increase in (F) behavioral choice of gambles following earlier non-rewarded gamble-arm trials ($F_4 = 6.940$, $p < 0.001$) results in (G) higher number of disadvantageous actions when evidence for the safe arm was high (EV < 1; $F_4 = 7.177$, $p < 0.001$).

(H) Optogenetic inactivation during reward experience on gamble (RR) or safe-arm trials (SafeR) or during the Run1 episode on any trial (R1) did not alter performance significantly. (E–G: data as mean ± SEM, $n_{control} = 22$, $n_{nor} = 17$, $n_{R1} = 11$, $n_{Rew} = 10$, $n_{SafeR} = 9$; one-way ANOVA, post hoc multiple comparison Tukey test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; data from 4–7 rats, see supplementary information and Figure S8)

Inactivation of neuronal firing, precisely timed to negative outcome, impeded behavioral performance in agreement with lesions or inactivation studies of the prelimbic cortex (Birrell and Brown, 2000; Marquis et al., 2007), indicating the direct involvement of the discovered firing patterns in value-based decision making. Our optogenetic perturbations provide evidence that animals escalate disadvantageous risk-seeking behavior based on a failure to adapt their decisions adequately following a non-rewarded gamble arm trial (Figures 7 and S8). We cannot attribute the behavioral changes solely to the inactivation of choice-predicting cells as a range of other neurons depict increased firing during non-rewarded trials. But the failure to adequately integrate and adapt to negative reinforcement for optimizing economic decisions follows observations on the importance of the prelimbic cortex in strategy, rule, and set-shifting tasks (Tervo et al., 2014; Durstewitz et al., 2010).

Choice-predicting cells increase their firing during negative outcome on the gamble arm only when the animal will go to the safe arm in the subsequent trial. Thus, they do not just signal negative and unsatisfactory outcome but much more act as a stable indicator when an imminent choice adaptation is favored. This presents a signal that provides a potent driver of behavioral change during reward evaluation independent of choice evidence and goal value. The increase in firing rate might influence goal value updating in interconnected networks to guide choice to an alternate goal in the future. This is a signal reminiscent of "regret," a long-standing concept of decision-making in economics (Bell, 1982; Loomes and Sudgen, 1982). Although regret has already been linked to dorso-medial and dorsolateral prefrontal cortex (PFC) in humans (Chua et al., 2009), a neuronal mechanism has not yet been identified. Regret, in contrast to disappointment, carries self-blame about one's choice and thus a stronger negative affective reaction to the outcome of the agent's choice. A better choice could have been made, which potentially carries a more direct effect on subsequent choices. Regret can only be experienced once the agents either can infer the likely outcome of alternative options or is informed about the outcome of the not-taken alternative. Our task design fulfills this requirement as, during the reward episode of non-rewarded gamble-arm trials, the animal can infer that it would have received a small reward on the safe-arm. Prelimbic regret-based signals may link reciprocal OFC and ventral striatal signals (Steiner and Redish, 2014) and potentially drive or reinforce decisions contrary to value but in favor of utility and complement utility-based decision processes specifically during decision making under uncertainty (Bell, 1982; Loomes and Sudgen, 1982).

In conclusion, the firing patterns of choice-predicting cells present an intrinsic evaluation process of negative outcomes that signal future choice and provide a neuronal framework for understanding individual decisions during economic choice. These signals' apparent independence from value representations and reward prediction error signals provides a nuanced view on neuronal correlates and models of decision-making processes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ Surgeries: Microdrive implantation, Virus injection and optical fiber implantation
  - ○ Behavioral Paradigm
  - ○ *In vivo* electrophysiology
  - ○ Anaesthesia experiments
  - ○ Optogenetic experiments
  - ○ Histology
  - ○ Analysis and statistics
  - ○ Behavioral Analysis
  - ○ Behavioral modeling
  - ○ Expected Value Model
  - ○ Multiple regression analysis
  - ○ Demixed principal component analysis
  - ○ Elastic-net regularization and general linear model prediction
  - ○ Multiple regression analysis for control of task variable influence over firing of choice-predicting cells
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and three tables and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.10.050.

ACKNOWLEDGMENTS

We thank Todor Asenov, Erzsebet Borok, and Romana Hauer for technical help, Stephane Ciocchi for scientific advice, and Jozsef Csicsvari, David Dupret, Karl Friston, and Mate Lengyel for comments on an earlier version of the manuscript. This work was supported by grant LS14-095 of the Vienna Science and Technology Fund, and the NIH (MH111529, MH071679) G.F. and J.D.

AUTHOR CONTRIBUTIONS

J.P. carried out the experiments; J.P. led the analysis of the paper in collaboration with N.M., H.M.-V., P.A., G.D., and T.K.; J.D. and G.F. developed and generated AAV2/1-mDlx-ChR2; J.P. and T.K. planned the experiments and wrote the paper with editorial input from all authors. T.K. supervised all aspects of the work.

DECLARATION OF INTERESTS

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: January 22, 2018
Revised: July 23, 2018
Accepted: October 29, 2018
Published: December 6, 2018

REFERENCES

Arlot, S., and Celisse, A. (2010). A survey of cross-validation procedures for model selection. Stat. Surv. *4*, 40–79.

Balleine, B.W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology *37*, 407–419.

Barack, D.L., Chang, S.W.C., and Platt, M.L. (2017). Posterior cingulate neurons dynamically signal decisions to disengage during foraging. Neuron *96*, 339–347.e5.

Bell, D.E. (1982). Regret in decision making under uncertainty. Oper. Res. *30*, 961–981.

Birrell, J.M., and Brown, V.J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. J. Neurosci. *20*, 4320–4324.

Botvinick, M.M., Cohen, J.D., and Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: An update. Trends Cogn. Sci. *8*, 539–546.

Burnham, K.P., and Anderson, D.R. (2002). Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach (Springer).

Chua, H.F., Gonzalez, R., Taylor, S.F., Welsh, R.C., and Liberzon, I. (2009). Decision-related loss: regret and disappointment. Neuroimage *47*, 2031–2040.

Cox, D.R. (1958). Two further applications of a model for binary regression. Biometrika *45*, 562–565, CR–169.

Csicsvari, J., Hirase, H., Czurko, A., and Buzsáki, G. (1998). Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. Neuron *21*, 179–189.

Dimidschstein, J., Chen, Q., Tremblay, R., Rogers, S.L., Saldi, G.A., Guo, L., Xu, Q., Liu, R., Lu, C., Chu, J., et al. (2016). A viral strategy for targeting and manipulating interneurons across vertebrate species. Nat. Neurosci. *19*, 1743–1749.

Dolan, R.J., and Dayan, P. (2013). Goals and habits in the brain. Neuron *80*, 312–325.

Durstewitz, D., Vittoz, N.M., Floresco, S.B., and Seamans, J.K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron *66*, 438–448.

Friedman, A., Homma, D., Gibb, L.G., Amemori, K.I., Rubin, S.J., Hood, A.S., Riad, M.H., and Graybiel, A.M. (2015). A corticostriatal path targeting striosomes controls decision-making under conflict. Cell *161*, 1320–1333.

Fujimoto, A., Tsurumi, K., Kawada, R., Murao, T., Takeuchi, H., Murai, T., and Takahashi, H. (2017). Deficit of state-dependent risk attitude modulation in gambling disorder. Transl. Psychiatry *7*, e1085.

Fujisawa, S., Amarasingham, A., Harrison, M.T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. Nat. Neurosci. *11*, 823–833.

Gardner, M.P.H., Conroy, J.S., Shaham, M.H., Styer, C.V., and Schoenbaum, G. (2017). Lateral orbitofrontal inactivation dissociates devaluation-sensitive behavior and economic choice. Neuron *96*, 1192–1203.e4.

Gershman, S.J. (2016). Empirical priors for reinforcement learning models. J. Math. Psychol. *71*, 1–6.

Glimcher, P.W., and Fehr, E. (2014). Neuroeconomics. Decision Making and the Brain (Academic Press).

Hare, T.A., Schultz, W., Camerer, C.F., O'Doherty, J.P., and Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. Proc. Natl. Acad. Sci. USA *108*, 18120–18125.

Hastie, T., Tibshirani, R., and Friedman, J. (2011). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer New York).

Hayden, B.Y., Nair, A.C., McCoy, A.N., and Platt, M.L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. Neuron *60*, 19–25.

Hayden, B.Y., Heilbronner, S.R., Pearson, J.M., and Platt, M.L. (2011). Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J. Neurosci. *31*, 4178–4187.

Hazan, L., Zugaro, M., and Buzsáki, G. (2006). Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. J. Neurosci. Methods *155*, 207–216.

Hunt, L.T., and Hayden, B.Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. Nat. Rev. Neurosci. *18*, 172–182.

Ito, H.T., Zhang, S.J., Witter, M.P., Moser, E.I., and Moser, M.B. (2015). A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. Nature *522*, 50–55.

Kable, J.W., and Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. Nat. Neurosci. *10*, 1625–1633.

Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., and Rushworth, M.F.S. (2006). Optimal decision making and the anterior cingulate cortex. Nat. Neurosci. *9*, 940–947.

Kepecs, A., Uchida, N., Zariwala, H.A., and Mainen, Z.F. (2008). Neural correlates, computation and behavioural impact of decision confidence. Nature *455*, 227–231.

Kerns, J.G., Cohen, J.D., MacDonald, A.W., 3rd, Cho, R.Y., Stenger, V.A., and Carter, C.S. (2004). Anterior cingulate conflict monitoring and adjustments in control. Science *303*, 1023–1026.

Kim, J.N., and Shadlen, M.N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat. Neurosci. *2*, 176–185.

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepecs, A., Mainen, Z.F., Qi, X.L., Romo, R., Uchida, N., and Machens, C.K. (2016). Demixed principal component analysis of neural population data. eLife *5*, 1–36.

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. Science *302*, 1181–1185.

Kolling, N., Wittmann, M., and Rushworth, M.F.S. (2014). Multiple neural mechanisms of decision making and their competition under changing risk pressure. Neuron *81*, 1190–1202.

Lindsay, A.J., Caracheo, B.F., Grewal, J.J.S., Leibovitz, D., and Seamans, J.K. (2018). How much does movement and location encoding impact prefrontal cortex activity? An algorithmic decoding approach in freely moving rats. Eneuro *5*, ENEURO.0023-18.2018.

Loomes, G., and Sugden, R. (1982). Regret theory: an alternative theory of rational choice under uncertainty. Econ. J. (Oxf.) *92*, 805–824.

Mansouri, F.A., Tanaka, K., and Buckley, M.J. (2009). Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. Nat. Rev. Neurosci. *10*, 141–152.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

Marquis, J.P., Killcross, S., and Haddon, J.E. (2007). Inactivation of the prelimbic, but not infralimbic, prefrontal cortex impairs the contextual control of response conflict in rats. Eur. J. Neurosci. *25*, 559–566.

Matsumoto, K., Suzuki, W., and Tanaka, K. (2003). Neuronal correlates of goal-based motor selection in the prefrontal cortex. Science *301*, 229–232.

McCoy, A.N., and Platt, M.L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. Nat. Neurosci. *8*, 1220–1227.

Ogawa, M., van der Meer, M.A., Esber, G.R., Cerri, D.H., Stalnaker, T.A., and Schoenbaum, G. (2013). Risk-responsive orbitofrontal neurons track acquired salience. Neuron *77*, 251–258.

Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model. Annu. Rev. Neurosci. *34*, 333–359.

Padoa-Schioppa, C., and Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. Nature *441*, 223–226.

Paxinos, G., and Watson, C. (2007). The Rat Brain in Stereotaxic Coordinates, Sixth Edition (Elsevier Academic Press).

Pribram, K.H., Wilson, W.A., Jr., and Connors, J. (1962). Effects of lesions of the medial forebrain on alternation behavior of rhesus monkeys. Exp. Neurol. *6*, 36–47.

Rangel, A., Camerer, C., and Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. *9*, 545–556.

Raposo, D., Kaufman, M.T., and Churchland, A.K. (2014). A category-free neural population supports evolving demands during decision-making. Nat. Neurosci. *17*, 1784–1792.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585–590.

Rushworth, M.F.S., Noonan, M.P., Boorman, E.D., Walton, M.E., and Behrens, T.E. (2011). Frontal cortex and reward-guided learning and decision-making. Neuron *70*, 1054–1069.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Shima, K., and Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. Science *282*, 1335–1338.

St Onge, J.R., and Floresco, S.B. (2010). Prefrontal cortical contribution to risk-based decision making. Cereb. Cortex *20*, 1816–1828.

Steiner, A.P., and Redish, A.D. (2014). Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. Nat. Neurosci. *17*, 995–1002.

Stauffer, W.R., Lak, A., and Schultz, W. (2014). Dopamine reward prediction error responses reflect marginal utility. Curr. Biol. *24*, 2491–2500.

Sugrue, L.P., Corrado, G.S., and Newsome, W.T. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. Nat. Rev. Neurosci. *6*, 363–375.

Sul, J.H., Kim, H., Huh, N., Lee, D., and Jung, M.W. (2010). Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron *66*, 449–460.

Sutton, R.S. (1992). Gain adaptation beats least squares. Proc, 7th Yale Work. 161–166.

Sutton, R.R.S., and Barto, A.G.A. (1998). Introduction to Reinforcement Learning (MIT Press Cambridge).

Svoboda, K., and Li, N. (2018). Neural mechanisms of movement planning: motor cortex and beyond. Curr. Opin. Neurobiol. *49*, 33–41.

Tervo, D.G.R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., and Karpova, A.Y. (2014). Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. Cell *159*, 21–32.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. R. Stat. Soc.: Ser. B Stat. Methodol. *73*, 273–282.

Tsutsui, K.I., Grabenhorst, F., Kobayashi, S., and Schultz, W. (2016). A dynamic code for economic object valuation in prefrontal cortex neurons. Nat. Commun *13*, 7:12554.

Tversky, A., and Kahneman, D. (1992). Advances in prospect-theory: cumulative representation of uncertainty. J. Risk Uncertain. *5*, 297–323.

Zeeb, F.D., Baarendse, P.J.J., Vanderschuren, L.J.M.J., and Winstanley, C.A. (2015). Inactivation of the prelimbic or infralimbic cortex impairs decision-making in the rat gambling task. Psychopharmacology (Berl.) *232*, 4481–4491.

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. J. R. Stat. Soc.: Ser. B. Stat. Methodol. *67*, 201–320.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
VGAT-GP-Af1000-1	Frontier Institute, Japan	https://www.frontier-institute.com/wp/antibodies/?lang=en
Anti-Chr2,monoclonal antibody Clone 15E2	mfd Diagnostic Germany	https://www.mfd-diagnostics.com/de/
Alexa488 anti-mouse	Jackson ImmunoResearch Laboratories	RRID:AB_2340850
Cy3 secondary Antibody	Jackson ImmunoResearch Laboratories	RRID:AB_2338000
Bacterial and Virus Strains		
AAV2/1-mDlx-ChR2	Dimidschstein et al., 2016	https://www.addgene.org/83898/
Experimental Models: Organisms/Strains		
Long-Evans male rats	Charles River Laboratories	https://www.criver.com/
Software and Algorithms		
MATLAB 2016 and 2017a	Mathworks	https://www.mathworks.com/products/matlab.html
SigmaPlot v.13	Systat	https://systatsoftware.com/
KlustaKwik	KlustaKwik	http://klustakwik.sourceforge.net/

CONTACT FOR REAGENT AND RESOURCE SHARING

Correspondence and requests for materials should be addressed to johannes.passecker@meduniwien.ac.at or thomas.klausberger@meduniwien.ac.at (Lead contact).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All data presented was obtained from 12 Long-Evans male rats (320-440 g 3 to 4 months at the time of surgery, Charles River Laboratories). Once surgeries were performed animals were housed individually in Plexiglas cages (42 × 27 × 30 cm). Rats were housed under a 12h light/12h dark cycle and all experiments occurred in the light phase. Until rats had recovered from surgeries, animals had *ad libitum* access to food. During behavioral experimental periods, rats received restricted amounts of food to reach 85% of the preoperative weight. They always had *ad libitum* access to water. All experimental procedures were performed under an approved license of the Austrian Ministry of Science and the Ethical Committee of the Medical University of Vienna.

METHOD DETAILS

Surgeries: Microdrive implantation, Virus injection and optical fiber implantation

Rats were anesthetized with a mix of oxygen and isoflurane (induction 5%, maintenance 2%). Animals were shaved and fixed via ear bars on a stereotaxic frame (Narishige). Body temperature was monitored throughout the surgery and stabilized via a heating pad. Local (xylocaine® 2%) and systemic analgesics (Metacam® 2mg/ml, 0.5ml/kg) were applied. Vita-Pos® was applied to protect the cornea and iodine solution was used to disinfect the surgery site. The skull and bregma were exposed and six stainless steel screws were anchored into the skull. In case of Microdrive implantation, the two posterior screws above the cerebellum served as ground and reference. Stereotaxic coordinates (Paxinos and Watson, 2007) are detailed in the Table S2.

Craniotomies were performed at the respective coordinates, the dura mater was removed and saline solution applied to avoid surface oedemas. To avoid dehydration of the animal, every two hours Ringer's solution (10 ml/kg) was administered subcutaneously. In the case of freely-moving electrophysiological recordings, animals were implanted with custom-made micro-drives (Miba Machine Shop, IST Austria) containing 15 independently moveable tetrodes made of four twisted tungsten micro-wires (12.7 μm inner diameter, California Fine Wire Company). To reduce impedance, tetrode tips were gold-plated to reach 100 - 500 kΩ. Once the wires of the micro-drive were lowered to the respective sites, paraffin wax was applied around the guide cannulas to protect the open brain and dura. In the case of virus injections and optical fiber implantation a recombinant AAV2/1-mDlx-ChR2 (Dimidschstein et al., 2016) (@ 1.1 10E+9 viral genomes/ul) was introduced targeting GABAergic neurons. The virus solution was loaded into a pulled borosilicate glass capillary and pressure-injected into the mPFC with a Picospritzer III® (Parker Hannifin Corporation). About 600-800 nL of virus solution was injected over 10 min/track along two tracks for each hemisphere. Thereafter an optic fiber was lowered into both hemispheres in-between and slightly above the respective virus injection tracks.

In all experimental cases, the constructs were cemented to the skull and screws, with bone cement (Refobacin®), and if necessary sutures were applied. Animals were given post-operative analgesia (Dipidolor® 60 mg diluted per 500 mL drinking water) and at least 7 days of recovery time. Once animals recovered fully after a minimum of 7 days, behavioral training started.

Behavioral Paradigm

All eight rats were trained on the gambling task and required on averaged 3 weeks of training to successfully integrate reward probabilities and magnitudes to optimize behavioral choices throughout the task. The wooden Y-maze was 55cm in height and the size of each arm was 80 cm x 11 cm. The maze was centrally placed in a darkened tent-like area. Animals were first habituated for three days to the experimenter, the room, and the baited maze. During the first phase of the behavioral training, animals received the same reward (2 × 20 mg, TestDiet) on the end of both goal arms with a high reward probability (90%) via automated pellet feeders (Camden Instruments Ltd). Once rats successfully ran toward the goal arms and were used to being manually placed on the home arm, differential reward probabilities for the gamble-arm were introduced. Here, as in the final version of the task, animals had to differentiate between a safe-arm, where they would receive always one 20mg reward, and a gamble-arm. On this gamble-arm, reward probability significantly changed between three blocks within one daily session. One block centered around 12.5% reward probability and thus favored safe-arm choices. The gamble-arm should be favored by the animals during the block where the animal receives 4 pellets of 20 mg around 75% of the times. The third block left the rat in a more ambiguous state as the reward probability centered around 25% indicating no clear preferable choice between the two arms. To help animals estimate reward probabilities 8 forced trials per goal arm accompanied probability changes. The path to the opposite starting arm was prevented by a plastic pot in forced trials. This was followed by 34 free-choice trials per block, where animals could freely choose between the two arms. In order to avoid arm biases or temporal learning patterns, gamble-arm and safe-arm, as well as block order, were randomly interchanged between days and rats (during training and testing). Additionally, to avoid pattern learning and repetitive behavioral choice-patterns, we regularly skipped forced-trial segments and thus provided also non-guided reward probability changes on the gamble-arm. Thus animals could not solely rely on reward outcomes during forced trials for an optimal choice strategy but had to actively track and integrate reward outcomes across the free-choice trials for reward maximization. The resulting choice adaptations in Figure 1E indicate a high degree of feedback integration during free-choice trials across all task segments. Once learned, a door was introduced at the beginning of the trial, which required the animals to wait for two seconds, to minimize repetitive behavioral patterns. Before recording started, animals had to distinctly favor the optimal arms in the two respective blocks and perform within all three blocks on three consecutive days. During electrophysiological recordings, the animal's position was tracked with either one or three (n = 20 sessions) LEDs of different colors detected at 20 frames per second by an overhead video camera (Sony). In all other experiments, the animal's position was tracked with only one LED. All behavioral experiments were performed in the same room and within the same environment. The maze was cleaned with an odour-free solution after each session for each animal. The arm-identity swap experiment was performed for three sessions. In two sessions the animal experienced 100 trials with a constant reward-probability of (35%) on the gamble-arm (incl. forced trials). After 50 trials, arm-identity was exchanged between the arms (without physically exchanging the arms). In the third session reward probability was not constant between the two behavioral blocks (cells were pooled from all three sessions).

In vivo electrophysiology

Tetrodes were lowered into the prelimbic cortex and moved before each recording day in order to record from new units. Unit signals were pre-amplified with a head-stage (HS-132A, 2 × 32 channels, Axona Ltd), and amplified 1000X via a 64-channel amplifier. A 64-channel converter computer card (Axona Ltd) was used to digitize at 24 kHz at 16 bit resolution. The signal was then down sampled offline to 20 kHz. To obtain single units, the signal was band-filtered (0.8 – 5 kHz) and spikes were defined by detecting signal amplitudes bigger than 5 SD above the mean. Each potential spike was sampled with 32 data points (1.6 ms) over 0.2 ms sliding windows and a principal component analysis was used to extract the first three components of the spike waveform (Csicsvari et al., 1998). KlustaKwik automatic clustering software (http://klustakwik.sourceforge.net/) was used to detect spike waveforms from putative neurons (Hazan et al., 2006). The clusters obtained were then manually verified by assessing the waveform shape, the modulation of waveform amplitude across tetrode channels, the temporal autocorrelation (to assess the refractory period of a single-unit) and cross-correlation (to assess a common refractory period across single-units). The stability of single-units was confirmed by examining spike features over time.

Anaesthesia experiments

Validation of optogenetic inhibition: The experiments (n = 2 rats, part of the behavioral inactivation experiments) were performed under Urethane (1.25 g/kg body weight) with additional doses of a ketamine/xylazine mixture (17 and 7 mg/ml, respectively; 0.02 - 0.1 ml). Once the dental cement was removed, optic fibers were slowly removed and custom-made acute opto-drives inserted into the prelimbic cortex. Drives consisted of 4 tetrodes surrounding a central optic fiber. Tetrodes protruded around 0.6mm from the optic fiber tip. The acute opto-drive was slowly advanced through the prelimbic cortex (1μm steps) with a stepper motor (Scientifica) and neuronal signals were recorded. Once optical effects were apparent, single-unit activity was recorded with simultaneous optical stimulation as used during the behavioral protocols (see below). Additionally, we performed 50Hz stimulations (280 times, 1ms every 19ms). Signals were transmitted to a RA16AC head-stage (Tucker-Davis Technologies) and to a 16-bit analog-to-digital converter (Cambridge Electronic Design), amplified through an EXT-16DX system (NPI Electronics) and recorded via Spike2 (Cambridge Electronic Design). Spike detection and single-unit isolation as described above. Validation experiments to assess the spread of optogenetic inhibition were performed in two additional animals (part of the behavioral inactivation experiments). Same procedures were generally applied as above with the difference that original optic fibers were kept in place and anteriorly tetrodes were introduced slowly with a stepper motor in a dorsoventral fashion through cingulate, prelimbic and infralimbic cortex. Recordings were performed in batches, allowing the brain to adjust at each step and depth was monitored.

Optogenetic experiments

Optogenetic silencing during the gambling task was achieved via a DPSS laser (IkeCool Corporation) generating blue light (473 nm) which connected to two bilaterally implanted optic fibers through a ferrule-sleeve system (Senko Ltd) with 20-30 mW output power delivered to the brain tissue for each individual optic fiber. We opted for an approx. 66Hz stimulation (280 times, 1ms stimulation every 14ms) triggered by the reward sensor on non-rewarded gamble-arm trials (Nor), rewarded gamble-arm trials (RR), rewarded safe-arm trials (SafeR) or triggered by the home sensor of the home arm (R1). We used the same optical stimulation protocol in the 2-chamber experiment. Here, the animal was placed into a square open maze (90x90x30cm) out of black painted wood. Two panels segregated the arena into two equally-sized compartments. During 8 minutes of exploration, animals chased pellets randomly delivered into the box via pellet feeders mounted above. The pellet feeders were activated and delivered at the exact same time to avoid behavioral biases (every 30 s.). Animals performed at least two consecutive sessions each. In each of the recording session, the laser stimulus was activated exclusively only on one side of the arena (counterbalanced between recordings) approx. every 15 s. (similar to the average trial time in the gambling task). The animal's position was tracked via a red LED mounted on the optical fiber cable. Two of those animals were used to confirm activation and inactivation of neuronal populations in the prelimbic cortex during urethane-induced anesthesia (as described above).

Histology

To confirm the position of the recording sites from freely-moving electrophysiological experiments, rats were then deeply anesthetized with urethane and lesions were made at the tip of the tetrodes applying a 30 μA unipolar current for 10 s (Stimulus Isolator, World Precision Instruments). Rats were perfused with saline followed by a 20 min fixation with 4% paraformaldehyde, 15% (v/v) saturated picric acid. Tetrodes were retracted; the micro-drive and the brain were sequentially extracted. Serial coronal sections were cut at 50 or 70 μm with a vibratome (Leica). Lesions were confirmed on an Olympus BX61 microscope. To detect the expression of ChR2 exclusively in GABAergic neurons, we incubated sections of interest in serum containing mouse anti-ChR2 monoclonal antibody (1/100, mfd Diagnostics) and vesicular GABA transporter (VGAT, Anti guinea-pig, 1/10000, Frontiers Institute, Japan) in 0.1M Tris-buffered saline containing, 1% normal horse serum and 0.1% Triton X-100. Sections were next incubated with Alexa488 anti-mouse (1/10000, Jackson Immuno Research Laboratories) and Cy3 (1/10000, Jackson Immuno Research Laboratories) fluorescent secondary antibodies. Immuno-histochemical analysis and image acquisition was performed on a confocal microscope (Leica TCS SP5).

Analysis and statistics

If not stated otherwise alpha is 0.05 and statistical testing was two-tailed each dataset was tested for normality (Kolmogorov-Smirnoff test). Additional statistical details for respective Figures can be found in Table S3.

Behavioral Analysis

Three pairs of sensors were placed on the maze. One pair indicated the start of the arm, at the beginning of the home arm, and one pair of sensors were located on each of the goal arms. A crossing of the reward sensors triggered either delivery or withholding of reward. In addition, sensor signals allowed us to separate the behavior into individual trials. We used the tracking data to detect the start of the run precisely (crossing of the door) and define the division point as a position along the maze where the first significant difference (in x/y position) between all trials within a session (t test, $p < 0.05$) occurred (Sul et al., 2010). This allowed us to reliably subdivide each trial into four behaviorally relevant episodes (run1, run2, reward, inter-trial). Run1 was defined as the episode between the home arm sensor activation and the division point. Run2 was defined as the period between the division point and the reward sensor. The reward episode was defined as the period between the reward sensor activation and the grabbing of the animal. The inter-trial period was defined as the return of the animal (between the grabbing and the return to the home arm sensor). For the calculation of the firing rate, each trial was subdivided into 30 time-bins (9 for run1, 3 for run2, 9 for reward and 9 for inter-trial) which was based on the average trial and episode times of the animals. In order to compare firing rates of single trials against each other and across sessions we opted for a combined spatial and time normalization. Within each episode and trial, the time was normalized for the respective bin number. We excluded single trials where the total trial time was above 2 standard deviations of the mean. The mean bin time was 485.6ms ± 139.3 (mean ± Stdev), resulting in an average trial time of 14.5 s. These 30 time-bins were subsequently used in all analysis. We focused our analysis exclusively on free-choice trials. To control for motoric confounds during our analysis of electrophysiological neuronal signals, we calculated maze trajectory (trk) for each arm and session (i) and calculated the bin-by-bin (b) variation to the mean trajectory values (mtrk) of the respective arm to allow direct comparison to the variance in firing rate during the same bins.

$$V_{bi} = \sqrt{(mtrk_{bix} - trk_{bix})^2 + (mtrk_{biy} - trk_{biy})^2}$$

In 20 sessions where he had access to 3LED tracking we calculated the relative change in angular heading to evaluate head-movement speed. In addition, we extracted head-direction (in degrees) for each time point and calculated mean-directional heading for each time-bin.

Behavioral modeling

A Rescorla-Wagner based reinforcement learning algorithm was applied to model the internal variables related to the decision process (Sutton and Barto, 1998) The subjective values of outcomes, called action values, are updated at each trial according to the difference between the expected and the actual reward in the following way:

If a rat goes to the safe-arm:

$$Q_s(t+1) = Q_s(t) + \alpha(R(t) - Q_s(t))$$

and

$$Q_g(t+1) = Q_g(t)$$

where $Q_s(t)$ is the action value of the safe-arm in trial t, $Q_g(t)$ is the action value of the gamble-arm in trial t, $R(t)$is the reward that the rat received in trial t (1 pellet for safe-arm and 0 or 4 pellets for the gamble-arm) and á is the learning parameter, indicating the amount of reward feedback used by the animals in their choice patterns. Alpha values close to 0 would indicate a minimal use of the reward feedback for action value updating. High alpha values indicate a high impact of the reward feedback on the animal's upcoming choice. The equation is analogous if the rat chooses a gamble-arm.

Initially, we assume that the rat does not have any preferences of the arms at the start of the block, so we define $Q_s(t)$ and $Q_g(t)$ to be zero for $t = 0$. Actions are chosen according to the soft-max selection criterion.

The probability of the rat going to the gamble-arm in trial t is:

$$P_g(t) = \frac{1}{1 + e^{-\beta\left(Q_g(t) - Q_s(t)\right)}}$$

where β is the inverse temperature parameter that defines the degree of exploration in the action selection. When β is close to 0, the contribution of the difference of the action values in the equation becomes smaller and the resulting behavior is more stochastic and explorative. When β is increased, the difference in selection probability becomes bigger and the animal's behavior is more exploitative. Assuming that the learning rate and the inverse temperature are not inherent personality features, but rather depend on learning state and motivation, they may change from session to session for each rat. Parameters α and β were estimated for each session separately using maximum likelihood estimation. We conducted a grid search for parameters α between 0 and 1 and for β between 0 and 50. For each point of the grid we calculated the likelihood function $f(x;\alpha,\beta)$, representing the likelihood that the we observe data x with a particular set of parameters and is defined as: $f(x;\alpha,\beta) = \prod_t P_{a(t)}(t)$, where $a(t)$ is action in trial t.

We optimized the parameters by maximizing the likelihood function of the observed and predicted choices of a rat in each session.

We used a cross-validation method for time series to estimate the predictive strength of the model. We train the data on $x_1, x_2, x_3, \ldots x_k$ data points and use the estimated parameters to predict the choice in the $(k+1)$-th trial. We performed this for the last 30 choice trials of each session and count the percentage of successful predictions (Table S1). To be able to compare different models we calculated the corrected Akaike Information Criterion as follows:

$$AIC_c = -2 * \ln\left(\max_{\alpha,\beta}(f(x;\alpha,\beta))\right) + 2 * K,$$

where K is the number of parameters used in the model (Burnham and Anderson, 2002).

In the first reinforcement model we included information from forced trials and is defined as follows:

$$P_g(t) = \begin{cases} \dfrac{1}{1 + e^{-\beta\left(Q_g(t) - Q_s(t)\right)}}, \text{if t is a free choice trial} \\ 1, \text{if t is a forced gamble trial} \\ 0, \text{if t is a forced safe trial.} \end{cases}$$

In the second reinforcement learning model, we included the reward ratio in the maximum likelihood estimation, modeling the perception of the higher reward on the gambling arm compared to the safe-arm. A logistic regression model (Cox, 1958) and a Win-stay lose-shift model (WSLS) were used as reference points for the reinforcement learning models' performance. The WSLS model predicts that the rat would pick the same arm choice after a choice was rewarded, and change its choice after an unrewarded trial. The logistic regression model is a discrete choice model that estimates probabilities of choosing one option or the other as predicted by past choices and past choice outcomes.

By assigning choice variables to be 1 for gamble and 0 for safe trials, defining 1 to k trials prior to trial t as $C_1(t),\ldots,C_k(t)$, and choice outcomes of 1 to k trials prior to trial t as $R_1(t),\ldots,R_k(t)$, we can formulate the probability of going to the gamble-arm as

$$P_g(t) = \frac{1}{1 + e^{-\left(\sum c_i C_i(t) + r_i R_i(t)\right)}},$$

where $c_1,\ldots,c_k$ and $r_1,\ldots,r_k$ are regression coefficients. The regression coefficients were estimated using the built-in MATLAB function mnrfit, based on maximum likelihood estimation. We performed the estimation for each session separately and evaluated the prediction accuracy for the current choice by incorporating choice and reward outcomes of up to 9 previous trials. In order to test whether a learning model with an adjustable learning rate would provide us significantly better prediction accuracy, we tested the K1 algorithm of the dynamic learning rate model described by Sutton (1992). Parameter estimations were performed via the built-in MATLAB function fmincon. We used a cross-validation technique as described above to obtain the number of successful predictions for each group of variables. We use the Akaike criterion to compare the different models for each session. Both the second reinforcement learning model as well as the dynamic learning rate model performed similarly good and significantly better than the remaining models (Table S1). We opted for the reinforcement learning model 2 in favor over the dynamic learning rate model as the latter did not significantly improve the performance. Importantly, results of choice-predicting cells were consistent for both models (see Figure S5I).

Expected Value Model

In order to see how an agent with perfect memory would act we fit an expected value model.

The model assumes that the rats are aware that the safe arm always brings 1 pellet and have therefore set the expected value of the safe arm in, $EV_{safe}(t)$, to be 1, regardless of trial *t*. The expected value of the gamble-arm at trial *t* is defined as

$$EV_{gamble}(t) = p_{gamble}(t) * 4^{\rho},$$

where $p_{gamble}(t)$ is the proportion of positive outcomes following a gamble-arm choice. In the EV model we assume that the animals are aware of the block structure and reset the history of outcomes at the start of each block. The parameter ρ defines the curvature of the value function (Tversky and Kahneman, 1992) and determines the risk preference of the rat. Higher ρ values indicate a higher risk-seeking prevalence of the animal's choice behavior.

The expected values are transformed into probabilities with the same choice model we use in the reinforcement learning models. We define the probability of choosing the gamble-arm on trial *t* as:

$$P_{gamble}(t) = \frac{1}{1 + e^{-\beta\left(EV_{gamble}(t) - EV_{safe}(t)\right)}},$$

where β represents the inverse temperature parameter as in other models.

The parameters β and ρ are estimated using the maximum likelihood estimation method, regularized with gamma priors on β (Gamma (4.83, 0.73); Gershman, 2016) and weakly informative normal priors on ρ (N(1, 0.3)). Disadvantageous choices were defined as any gamble-arm choice below an EV of 1 and any safe-arm choice above an EV of 1.

Multiple regression analysis

A semi-automated multi-regression analysis for each time-bin and for each cell was calculated estimate the correlation between the firing rate of the cells and the behavioral variables of reward, goal arm and choice evidence (Figure 2B). To check whether the variables contribute to the firing rate variance of our recorded neurons we used the following model:

$$FR_i(t) = \beta_0 + \beta_1 * A(t) + \beta_2 * R(t) + \beta_3 * CE(t) + \varepsilon(t),$$

where $FR_i(t)$ is the firing rate in trial t at each certain time-bin i, $A(t)$ is the chosen goal arm, $R(t)$ is the reward in the gamble trials, CE is the choice- evidence denoted as the probability value to choose the gamble-arm, $\varepsilon(t)$ is the error term and β_0, β_1, β_2, β_3, are the regression coefficients. We used a cross validation (CV) to select the most appropriate subset of variables for the model (Arlot and Celisse, 2010). For each subset of the k variables we found the matrix of coefficients $\hat{a}$ that solves the regression equation and calculated the cross validation statistic (E_{cv}) in the following way:

$$E_{cv} = \frac{1}{n_{trials}} * \sum_{i=1}^{n} \left(\frac{e_i}{1 - h_i}\right)^2,$$

where n_{trials} is the number of trials and h_i are the diagonal terms of the matrix $H = \hat{a}(\hat{a}'\hat{a})^{-1}\hat{a}'$. The subset of variables with the smallest CV was then used in the final model together. Once the variables have been selected we included all the possible interactions in the model and used only the ones that improve the CV statistic. The variables and interactions chosen were then used in the final model to estimate the regression coefficients and p values. We plot the fractions of neurons with firing patterns that significantly correlate with the arm-related variable, reward variable and the choice evidence variable (Figure 2B). A neuron was only counted if three consecutive time bins were significantly correlated ($p < 0.05$). Thus No-reward activated cells had to demonstrate a significant correlation with the absence of reward on the gamble-arm for three consecutive time bins (on average 1350 ms) during the reward episode.

Demixed principal component analysis

In order to confirm and validate our data and specifically observe the contribution of the task parameters on the neuronal firing rate we opted for a demixed principal component approach (DPCA) (Kobak et al., 2016). Standard principal component analysis finds the optimal orthogonal components which explain most of the variance in the data. However, these components are a mixture of the task variables and, even though the variability of the data is highly explained, this explanation does not relate well to the initial parameters. The DPCA finds the components which capture the maximum amount of variance in the data while maintaining the task parameters as segregated as possible. In general, we followed the procedure presented by Kobak et al., 2016 for sequentially recorded data and used the provided MATLAB scripts to analyze the dataset (https://github.com/machenslab/dPCA). The firing rates per cell and time bins for all gamble-arm trials were exported into the respective multi-dimensional input matrices. We defined the three choice evidence periods as stimulus variables and reward occurrence on the gamble-arm as the decision variable. Only cells with a firing rate of less than 50Hz were accepted. A minimum of 2 trials for each combination of variables was required for a session to be accepted. Specifically two rewarded gamble-arm trials were unlikely to occur when choice evidence for a gamble was low. Hence, 380 neurons were used as an input. We present the demixed principal component analysis with regularization and used cross-validation to measure time-dependent classification accuracy and a shuffling procedure (100x) to assess whether it was significantly above chance same as in[20] to establish the significance of individual components (Figure 2A).

Elastic-net regularization and general linear model prediction

In order to predict future choices (t+1) based on the firing rate of the current trial (t) we used the trial averaged firing rate of cells as predictors and future choices (t+1) as the response variable. If not stated otherwise, trial averaged firing rates were taken into account for the prediction models. Either we tested all non-rewarded trials (t) and asked which arm the animal will choose in the future trial (t+1) (Figures 4, 5, and S5) or we tested any trial to assess future choice (Figure S3A–S3C). On average, 22.3 neurons were successfully isolated in each recording session. Only sessions with a sufficiently high number of no-reward trials (at least 20 in total, and a minimum of 5 trials per condition) and at least 3 no-reward activated cells present were taken into account. This reduced the numbers of sessions tested for choice prediction after non-rewarded trials to 37 sessions (out of 45). For those 37, approx. 10.5 neurons per session were identified as no-reward activated cells and those sessions included on average 30.5 no-reward trials (median: 29). To reduce the number of features we applied an elastic-net regularization with a leave-one-out cross-validation to identify the most relevant predictors (Lasso/Elastic-net MATLAB in-built function based on (Zou and Hastie, 2005; Tibshirani, 2011; Hastie et al., 2011):

$$\min_{\beta_0\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(y_i-\beta_0-x_i^T\beta\right)^2+\lambda P_\alpha(\beta)\right)$$

where

$$P_\alpha(\beta)=\frac{(1-\alpha)}{2}\|\beta\|_2^2+\alpha\|\beta\|_1=\sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2+\alpha\left|\beta_j\right|\right)$$

and N is the number of observations; y_i the response of the observation i. x_i is a vector of p values at the observation i. λ is a positive regularization parameter. The parameters $\beta 0$ and β represent scalar and p-vector respectively. As λ increases, the number of nonzero elements of β decreases. The Elastic-net regularization sets more coefficients to zero with an increasing penalty term. We interpolate between the L^1-norm of β and the squared L^2 of β via the penalty term $P\alpha(\beta)$. Although, minimal differences (in cells selected) could be observed when different alpha (α) values (between 0 and 1, data not shown) were tested, the data presented here are based on an α of 0.2.

Once features were selected as input predictors (X_i) by the regularization for each session separately, we tested those on a general linear prediction model. When we tested for all future choices, 10% of the trials were withheld for the test-data. In all other instances, one trial was kept as test-data, and the remaining trials (training data) were used to generate a model with a leave-one-out cross-validation. The resulting model was tested on the remaining unseen test-data.

$$\mu=\beta_0+\sum\beta_iX_i+\varepsilon$$

Whereas β_0 is the intercept and ε is an error term. The procedure was repeated 30 times (average number of non-rewarded trials). The resulting average model accuracy per session was presented as prediction accuracy (Figures 4B, inlet, and S3A). Receiver operating characteristics were calculated based on the validation sets for each session (Figures 4B, S2F, S3B, and S5L). The accuracy of the prediction could not be attributed to a behaviorally increased likelihood of change after a non-rewarded trial. In fact, in 40% of non-rewarded trials animals did change their choice on the next trial. Regression coefficients presented in all figures are the averaged model coefficients per cell for each model repetition (Figures 4D and S3C).

In order to investigate change-predicting cells, which are reliable predictors across all choice evidence periods, and to negate the effects of negative feedback scaling we used the first derivative of the firing rate. The mean firing rate of trial t_{-1} (excluding the reward episode) was subtracted from the firing rate during the reward episode of trial t and used as input predictor for future choice in trial (t_{+1}) based on the firing rate within that episode. The same prediction model was then used as described above.

All calculations were made in MATLAB® (version 2016 and 2017a) and statistical analysis was performed in either MATLAB® or with SigmaPlot® (version 13).

Multiple regression analysis for control of task variable influence over firing of choice-predicting cells

This method allowed us to test whether choice-predicting signals persist to be identified once we control for variance in a range of task-parameters. Only those sessions were tested where angular head-directional data was available (sessions recorded with three tracking LEDs). Formula for regression:

$$\mathrm{FR_i}(t) = \beta_0 + \beta_1 * \mathrm{Q_g}(t) + \beta_2 * \mathrm{RPE}(t) + \beta_3 * \mathrm{CE}(t) + \beta_4 * \mathrm{Ang}(t) + \beta_5 * \mathrm{Trk}(t) + \varepsilon(t),$$

Where â are the respective coefficients, and FR the firing rate in trial t; CE: choice-evidence, RPE: Reward prediction error; Ang: changes in head-direction; Trk: changes in deviation from mean tracking and Q_g is the modeled action value of the Gamble arm for trial t. Instead of using firing rates as inputs into our elastic-net regression (as above) we use here the residuals for each individual neuron for the regression.

DATA AND SOFTWARE AVAILABILITY

The data and code that support the findings of this study are available from the corresponding authors upon request.