

<https://doi.org/10.1038/s42003-025-08729-x>

# Identification of conserved frontal neurophysiological markers of cognitive flexibility in humans and rats



Andre Der-Avakian<sup>1,9</sup>, Samuel A. Barnes<sup>1,9</sup>, Ty Lees<sup>1,9</sup>, Hans S. Schroder<sup>2</sup>, Brian D. Kangas<sup>1,2</sup>, Samantha R. Linton<sup>2</sup>, Stefanie Nickels<sup>2</sup>, Mykel A. Robble<sup>2</sup>, Micah Breiger<sup>2</sup>, Ann M. Iturra-Mena<sup>2</sup>, Rachel Lobien<sup>2</sup>, Sarah Perlo<sup>2</sup>, Emilia F. Cárdenas<sup>2</sup>, Genevieve P. Nowicki<sup>2</sup>, Zeyun Wu<sup>3</sup>, Hongyi Pan<sup>3</sup>, Daniel G. Dillon<sup>2</sup>, James P. Kesby<sup>4,5</sup>, Jack Bergman<sup>2</sup>, William A. Carlezon Jr<sup>2</sup>, Victoria B. Risbrough<sup>1,6</sup>, Eran Mukamel<sup>3</sup>, Stefan Leutgeb<sup>7</sup> & Diego A. Pizzagalli<sup>2,8</sup> ✉

Cognitive flexibility broadly describes behavioral alterations made in response to environmental changes and is fundamental for survival. While human and non-human animal assessments of cognitive flexibility are available, a systematic cross-species comparison of behavioral, neurophysiological, and computational markers of cognitive flexibility has not been reported. Using versions of a probabilistic reversal learning task aligned between humans and rats, electroencephalogram recordings reveal a frontal reward positivity (RewP) associated with unexpected reward outcomes. Reinforcement Q-learning models of both species' task behavior reveal that prediction error (PE) magnitude was significantly related to RewP amplitude. The stimulant drug modafinil alters PEs in rats without affecting the RewP in either species. These findings reveal analogous neurophysiological markers associated with PEs in humans and rats using equivalent tasks and identical computational analyses. This translational approach may improve the predictive validity of tests for novel pharmacotherapies and accelerate neuropsychiatric treatment by assessing neural mechanisms conserved across species.

Cognitive flexibility broadly refers to the ability to alter behavior in response to a changing environment<sup>1–3</sup>, and such adjustments are a vital component of navigating everyday life<sup>4,5</sup>. It describes the balance between repeating behaviors that yield beneficial outcomes toward a particular goal and modifying behaviors in response to environmental changes<sup>3,6</sup>. This modification of behavioral strategy is key to cognitive flexibility and enables previously learned rules to be updated to maximize rewarded outcomes. Cognitive flexibility has been studied across several key domains, including set-shifting, working memory, and reversal learning<sup>6</sup>. Moreover, human research has found that cognitive flexibility is associated with resilience to stress and negative life events<sup>4,7</sup> and overall higher quality of life<sup>8</sup>. Conversely, deficits in cognitive flexibility have been documented in a multitude of psychiatric disorders, including mood disorders<sup>9–11</sup>, schizophrenia<sup>12,13</sup>, obsessive

compulsive disorder<sup>14,15</sup>, and substance use disorder<sup>16</sup>. If the neurobiological processes underlying cognitive flexibility were better understood, they could be targeted for treatment in these patient populations.

Reversal learning tasks, particularly probabilistic versions, are ideal for assessing cognitive flexibility<sup>17–19</sup>. Although several variations of the probabilistic reversal learning (PRL) task exist, they are all based on a fundamental reinforcement learning paradigm whereby provided feedback is not always informative and the stimulus associated with the higher probability of reward is subject to change. For example, in a common PRL procedure, subjects are presented with two distinctive stimuli, one of which is associated with a high (e.g., 80%) probability of reinforcement, while the other is associated with a low (e.g., 20%) probability of reinforcement. By sampling both stimuli, subjects learn to maximize responding to the stimulus

<sup>1</sup>Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. <sup>2</sup>Department of Psychiatry, McLean Hospital & Harvard Medical School, Belmont, MA, USA. <sup>3</sup>Department of Cognitive Science, University of California San Diego, La Jolla, CA, USA. <sup>4</sup>Queensland Centre for Mental Health Research, Wacol, QLD, Australia. <sup>5</sup>Queensland Brain Institute, The University of Queensland, St. Lucia, QLD, Australia. <sup>6</sup>Veterans Affairs Center of Excellence for Stress and Mental Health, La Jolla, CA, USA. <sup>7</sup>Neurobiology Section and Center for Neural Circuits and Behavior, University of California San Diego, La Jolla, CA, USA. <sup>8</sup>Noel Drury, M.D. Institute for Translational Depression Discoveries, University of California, Irvine, CA, USA. <sup>9</sup>These authors contributed equally: Andre Der-Avakian, Samuel A. Barnes, Ty Lees. ✉e-mail: [dpizzaga@hs.uci.edu](mailto:dpizzaga@hs.uci.edu)

associated with a greater probability of reinforcement (i.e., target stimulus). After a pre-determined criterion (e.g., consecutive responses for the target stimulus), the reinforcement contingencies are reversed, and subjects need to identify the new target stimulus (identified as a reversal). Thus, PRL tasks require subjects to ignore occasional negative outcomes while responding to the target stimulus and avoid perseverating on a previously successful action when the reward contingencies are reversed. Individuals with major depressive disorder are sensitive to negative events, and for example, may respond to the occasional misleading negative feedback following a target response by prematurely responding for the other (i.e., non-target) stimulus<sup>20</sup>, whereas patients with obsessive compulsive disorder may perseverate on a stimulus that is no longer the target stimulus, despite the new relatively low reinforcement probability<sup>21</sup>. In either case, patients may complete fewer reversals, indicative of impairment in adaptive behavior.

Expression of flexible behavior during a PRL task requires successful and rapid reinforcement learning<sup>22</sup>, in particular, the ability to parse the likelihood of reward and identify the target stimulus. As a result, a violation of the expected outcome, such as the omission of reward following a target response (i.e., due to probabilistic feedback), should elicit a reward prediction error (PE). Frequent, but not occasional, PEs should facilitate a change in response strategy. PEs have been associated with changes in event-related potentials (ERPs) over frontocentral electrode sites, namely the Reward positivity (RewP)<sup>23,24</sup>. Originally named the feedback-related negativity<sup>25,26</sup>, and identified as a negative deflection following unexpected negative feedback (i.e., a negative PE) relative to expected positive feedback<sup>27,28</sup>, more recent evidence clarified that the RewP is driven by the response to reward feedback rather than non-rewarded outcomes<sup>23,29,30</sup>. Evidence suggests that the FRN/RewP originates in the anterior cingulate cortex (ACC)<sup>31–33</sup>. Alterations in the activity of midbrain dopamine neurons are known to code PEs<sup>34–36</sup>, whereby negative PEs elicit decreased firing of these neurons and phasic activation encodes positive PEs. This dopaminergic signal projects to the striatum and ACC and may modulate the FRN/RewP, which can subsequently predict behavioral adjustments<sup>26,37,38</sup>.

Non-human animal versions of the PRL task are critical for advancing our understanding of cognitive flexibility. In particular, the operant nature of the task enables translation across species; moreover, task behavior is well-characterized by reinforcement learning models, enabling application of the same (or very similar) models to data across species<sup>39</sup>. To study this phenomenon in non-human animals, rodent versions of the PRL task have been developed<sup>40,41</sup>. While human and non-human versions of the task are conceptually similar and may yield comparable behavioral results, to our knowledge, a systematic cross-species comparison of behavioral, computational, and neurophysiological markers of cognitive flexibility has not been reported. Such a comparison would be useful for drug development because putative treatments that demonstrate therapeutic behavioral effects and target engagement in non-human animals could be used in parallel human testing to accelerate drug discovery.

Thus, we modified human and rodent versions of a PRL task to align several parameters, enabling the comparison of behavioral performance across species. We measured the electroencephalogram (EEG) during the task to determine whether unexpected rewards were associated with a frontal RewP. Additionally, we applied a reinforcement Q-learning model to model behavior and linked it to EEG data from both species to compare the relationship between action values and frontal neurophysiological signals. Lastly, we administered comparable doses of modafinil, an indirect dopamine (DA) agonist that also has norepinephrinergic effects, to both species to determine whether modulation of dopaminergic signaling similarly altered the behavioral and neurophysiological indices of cognitive flexibility in humans and rats. Prior studies have shown that DA agonists increase reward-related ERPs during reward learning<sup>42</sup> and modafinil increases DA signaling in the striatum (via the inhibition of DA transporters), the neural substrate for error prediction<sup>43</sup>. Critically, we identified behavioral, computational, and neurophysiological markers of cognitive flexibility that were similar across species, providing a robust platform that could hasten treatment development.

## Results

We instructed humans ( $n = 54$ ) and trained rats ( $n = 11$ ) to perform functionally identical versions of a PRL task (Supplementary Fig. 1) that consisted of 300 trials and reinforced responses probabilistically using an 80%/20% schedule (see “Methods” section). For both species, the reward contingencies reversed after eight consecutive correct responses, and EEG recordings were obtained during task performance (see “Methods” section).

### Behavioral indices of cognitive flexibility and reinforcement metrics are consistent across humans and rats

During a single test session, humans completed  $4.4 \pm 0.23$  (mean  $\pm$  SEM) reversals per 100 trials (Fig. 1A), whereas rats completed  $2.6 \pm 0.40$  (Fig. 1B). Repeating a target response after reward delivery (i.e., target win-stay) and abandoning the target response after non-reward delivery (i.e., target lose-shift) reflect responsiveness to positive and negative feedback, respectively. In both humans (Fig. 1A) and rats (Fig. 1B), target win-stay probability was significantly greater than target lose-shift probability [humans:  $t(53) = 13.93$ ,  $p < 0.001$ ; rats:  $t(10) = 6.20$ ,  $p < 0.001$ ]. Increased target win-stay and reduced target lose-shift responding facilitated more reversals, as reflected by significant positive correlations between target win-stay probability and reversals in humans (Pearson  $r(52) = 0.44$ ,  $p < 0.001$ ; Fig. 1C) and rats ( $r(9) = 0.79$ ,  $p = 0.004$ ; Fig. 1D), and significant negative correlations between target lose-shift probability and reversals in humans ( $r(52) = -0.58$ ,  $p < 0.001$ ; Fig. 1C) and rats ( $r(9) = -0.73$ ,  $p = 0.012$ ; Fig. 1D). Thus, in both species, greater sustained responding for reinforced target responses, despite occasional misleading feedback, was associated with better task performance.

### Reinforcement learning models identify consistent patterns of behavioral responding across species

To gain a deeper insight into the behavioral mechanisms underlying PRL performance, we fit several Q-learning models to behavior (see Supplementary Methods). Consistent with our previous work<sup>41,44</sup>, we found that the model consisting of three free parameters (learning rate, inverse temperature, and forget parameters) best fit PRL performance (Supplementary Table 1). Moreover, parameter recovery and posterior predictive checks (Supplementary Fig. 2) demonstrated recoverable parameter estimates and alignment between simulated and observed PRL performance.

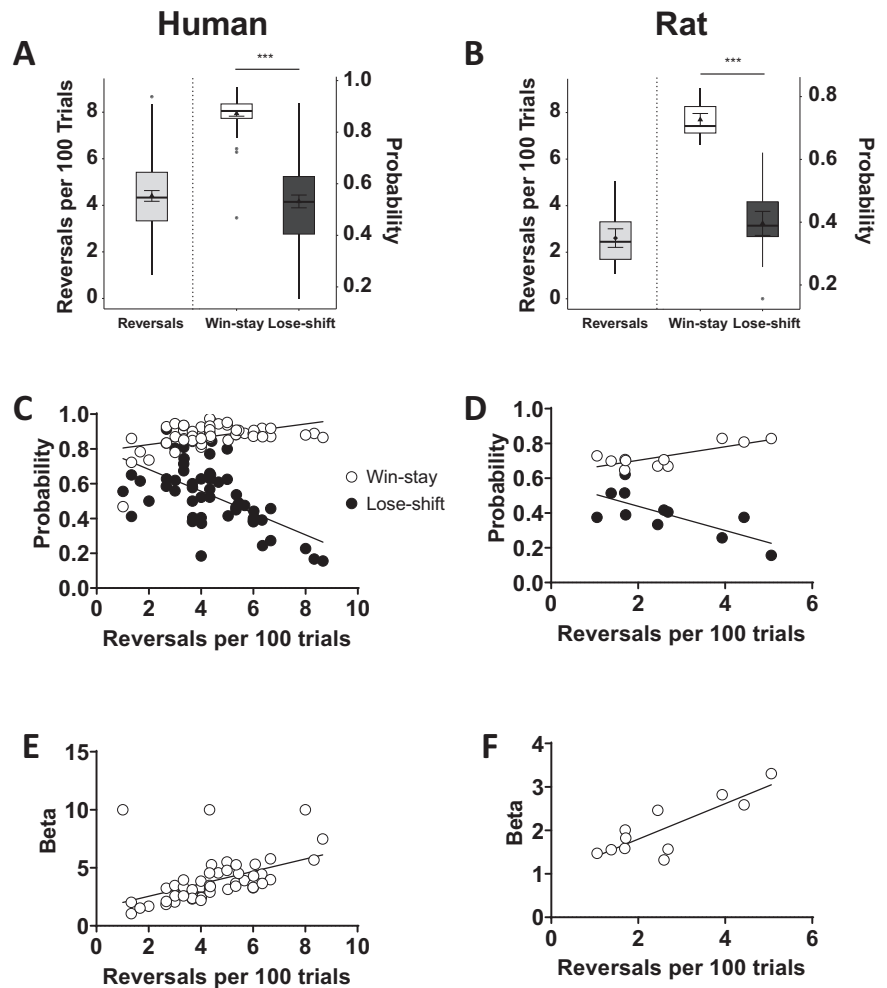
As subjects perform the PRL task, the value of each chosen action (i.e., Q-value) is updated based on reinforcement. Thus, during the task, the value of the target stimulus fluctuates to reflect the changing probability of reward delivery (see Supplementary Fig. 3A, B for representative Q-values across a test session for both species). As expected, humans assigned greater Q-values to target stimuli ( $0.710 \pm 0.007$ ) than non-target stimuli ( $0.329 \pm 0.01$ ) [ $t(106) = 25.51$ ;  $p < 0.001$ ] (Supplementary Fig. 3C), as did the rats ( $0.550 \pm 0.036$  vs.  $0.275 \pm 0.034$ ) [ $t(20) = 6.08$ ,  $p < 0.001$ ] (Supplementary Fig. 3C), indicating that both species learned to appropriately value stimuli based on experience throughout the task.

The beta parameter reflects the degree to which subjects explore both actions (lower beta value) vs. exploit the highest value action (higher beta value)<sup>45</sup>. Higher beta values were significantly associated with more reversals in humans ( $r(52) = 0.65$ ,  $p < 0.001$ ; Fig. 1E) and rats ( $r(9) = 0.69$ ,  $p = 0.019$ ; Fig. 1F). Thus, consistent with the target win-stay and lose-shift correlations described above, exploiting high value actions and limiting exploration to periods when those actions become less favorable (i.e., after reversals) was an adaptive strategy for both species. Conversely, there was no association between alpha or forget parameters and reversals in either species (Supplementary Fig. 4).

### Electrophysiological markers associated with reward expectancy are consistent across humans and rats

As humans and rats performed the PRL task, continuous EEG was recorded and averaged across rewarded and non-rewarded trials. A RewP emerged in frontal recording sites (e.g., frontocentral electrode (FCz) in humans (Fig. 2A); ACC in rats (Fig. 2B)) in response to positive vs. negative feedback. Highlighting spatial specificity, these effects were not present at parietal

**Fig. 1 | PRL behavior was comparable between humans (left) and rats (right) performing similar versions of the task.** Although humans ( $n = 54$ ; A) generally completed more reversals than rats ( $n = 11$ ; B), the probability of repeating a previously rewarded target response (win-stay) was greater than the probability of abandoning a previously unrewarded target response (lose-shift) in both species. Target win-stay probability was positively associated with the number of completed reversals in both humans (C) and rats (D), while target lose-shift probability was negatively associated with reversals in both species. Greater beta values, indicating a greater likelihood to exploit known information to maximize rewards, positively correlated with the number of completed reversals in humans (E) and rats (F). Reversal data are presented using box plots marking the median value (center line), the 25th and 75th percentiles (the outer edges), and  $\pm 1.5$  times the interquartile range (the whiskers), as well as the mean (triangle)  $\pm$  standard error of the mean (error bars).



electrodes in either species ( $p$ 's  $> 0.168$ ). The RewP peaked in the human FCz electrode approximately 200 ms following feedback and approximately 100 ms after reward feedback in the rat ACC, and reflected a more positive frontal signal for a rewarded vs. non-rewarded response (consistent with a positive PE). When broken down further into the four possible trial types, the average RewP amplitude was greater for rewarded compared to unrewarded trials after both target and non-target responses in both humans (Fig. 2C) and rats (Fig. 2D). In humans, a 2-way ANOVA revealed a significant Response  $\times$  Feedback interaction [ $F(1,33) = 6.10$ ,  $p = 0.019$ ], and Bonferroni post-hoc tests revealed significantly greater voltage during rewarded vs. unrewarded target trials ( $p = 0.002$ ). Moreover, a 1-way ANOVA revealed a significant linear effect of trial type [ $F(3,99) = 30.23$ ,  $p < 0.001$ ], and Bonferroni post-hoc tests revealed progressively increasing voltage that scaled with expected PE magnitude: Target/No Reward (expected to elicit the largest negative PE), Non-target/No Reward, Target/Reward, Non-target/Reward (expected to elicit the largest positive PE). In rats, a 2-way ANOVA did not reveal a significant Response  $\times$  Feedback interaction, but did reveal a significant main effect of Feedback [ $F(1,9) = 37.73$ ;  $p < 0.001$ ], and Bonferroni post-hoc tests revealed significantly more positive voltage during rewarded vs. unrewarded target trials ( $p < 0.001$ ). Critically, like humans, a 1-way ANOVA in rodents revealed a significant linear effect of trial type [ $F(3,27) = 18.52$ ,  $p < 0.001$ ], and Bonferroni post-hoc tests revealed progressively increasing voltage in the same order as humans.

#### Prediction errors associated with reward expectancy are consistent across humans and rats

Based on the value assigned to the chosen option on any trial, a PE can be computed depending on the outcome of that trial. For example, rewarded

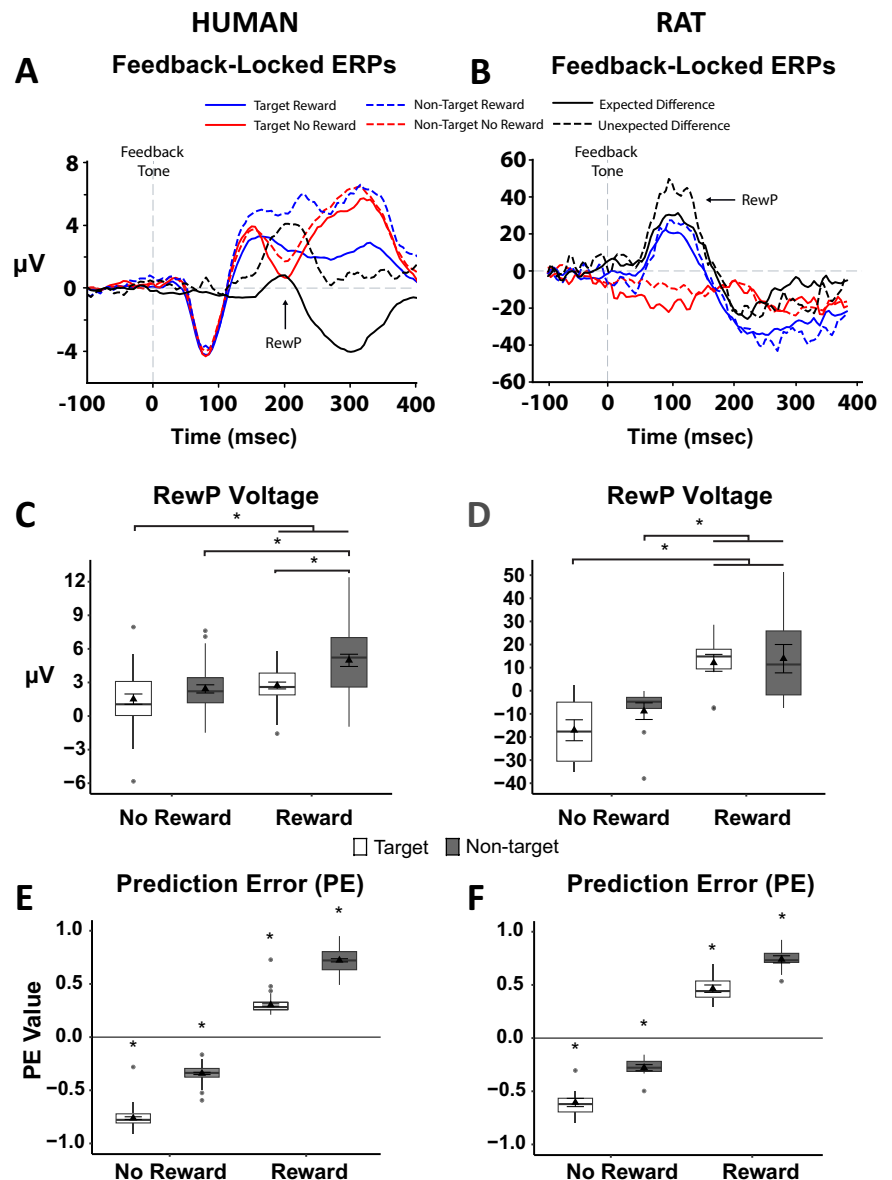
target responses (i.e., informative feedback) should elicit small positive PEs, whereas non-rewarded target responses (i.e., misleading feedback) should elicit large negative PEs<sup>46,47</sup>. Indeed, the PEs for rewarded and unrewarded outcomes were positive and negative, respectively, and PE magnitude was greater when the outcome of the selected action was unexpected (Fig. 2E, F). In humans, a 2-way ANOVA revealed a significant main effect of Feedback [ $F(1, 53) = 22,570.94$ ,  $p < 0.001$ ], but no Response  $\times$  Feedback interaction. Post-hoc tests revealed significantly greater PEs during rewarded vs. unrewarded target trials ( $p < 0.001$ ). Consistent with the linear effect of the four trial types on RewP voltage, 1-way ANOVA revealed a significant linear effect of trial type [ $F(3,159) = 3571.61$ ,  $p < 0.001$ ], and Bonferroni post-hoc tests revealed progressively increasing PE values in the same order as described for RewP voltage. In rats, a 2-way ANOVA revealed a significant main effect of Feedback [ $F(1, 10) = 5698.12$ ,  $p < 0.001$ ], but no Response  $\times$  Feedback interaction. Post-hoc tests revealed significantly greater PE during rewarded vs. unrewarded target trials ( $p < 0.001$ ). Similar to humans, the 1-way ANOVA across all four trial types in rodents revealed a significant effect [ $F(3,30) = 882.49$ ,  $p < 0.001$ ], due to progressively increasing PEs. Interestingly, in both species, PEs were more negative for unrewarded target vs. non-target responses and more positive for rewarded non-target vs. target responses (all  $p$ 's  $< 0.001$ ), suggesting that reward expectancy was greater following target vs. non-target responses in both groups.

#### Relationships between behavioral and electrophysiological measures of PEs

Next, we investigated associations between model parameters and electrophysiological markers. Using linear regression, we used trial-level data to

**Fig. 2 | Frontal feedback-locked event-related potentials (ERPs) and prediction error (PE) values were comparable between humans (left) and rats (right) during performance of the PRL task.**

Frontal (FCz in humans, anterior cingulate cortex in rats) ERPs were more positive during rewarded vs. unrewarded target responses in both humans (A) and rats (B), indicative of the reward positivity (RewP). The pattern of voltage changes during the RewP period across all four trial types was similar between humans ( $n = 34$ ; C) and rats ( $n = 11$ ; D). A nearly identical pattern emerged with regard to PE values in both humans ( $n = 34$ ; E) and rats ( $n = 11$ ; F). ERP traces are presented for all trial types (rewarded target = blue solid line; non-rewarded target = red solid line; rewarded non-target = blue dotted line; non-rewarded non-target = red dotted line) and expectancy-based difference waveforms (Expected = Rewarded Target – Non-rewarded Non-Target, and Unexpected = Rewarded Non-Target – Non-rewarded Target) and were averaged across all subjects and triggered to the tone predicting the presentation or omission of reward. Voltage and PE data are presented using box plots marking the median value (center line), the 25th and 75th percentiles (the outer edges), and  $\pm 1.5$  times the interquartile range (the whiskers), as well as the mean (triangle)  $\pm$  standard error of the mean (error bars).



determine the association between the ERP/LFP voltage at each timepoint within each trial and the PE associated with that trial. Notably, in both species (Fig. 3A, B), the regression coefficient peaked in the time window corresponding to the RewP. This indicates that during the RewP period, greater ERP/LFP amplitudes are associated with more positive PEs, whereas lower ERP/LFP amplitudes are associated with less positive PEs. That is, the more positive ERP/LFP for rewarded vs. unrewarded non-target responses appears to reflect the positive PE that occurs when rewards are unexpected. Analogous models using outcome and expected value (i.e., the sub-components of PE) are presented in Supplementary Fig. 5.

To provide additional validation for this regression approach, we used the intercept and regression coefficient generated at each timepoint to predict the ERP/LFP amplitude associated with various PEs. The predicted ERP/LFP traces (Fig. 3C, D) replicated the traces observed in humans and rats (Fig. 2A, B), and, importantly, differences in the predicted traces emerged at the time window corresponding to the RewP. Moreover, the predicted ERP/LFP was greater when the hypothetical PE was more positive. To ensure that the predicted ERP/LFP traces aligned with the activity we recorded from our human or rodent subjects, we plotted the ERP/LFP for trials associated with a PE greater than 0.5 or less than  $-0.5$  (Fig. 3E, F). As expected, and consistent with our predicted ERP/LFP values, more positive

PEs were associated with a more positive ERP/LFP, and this difference was most evident in the RewP time window.

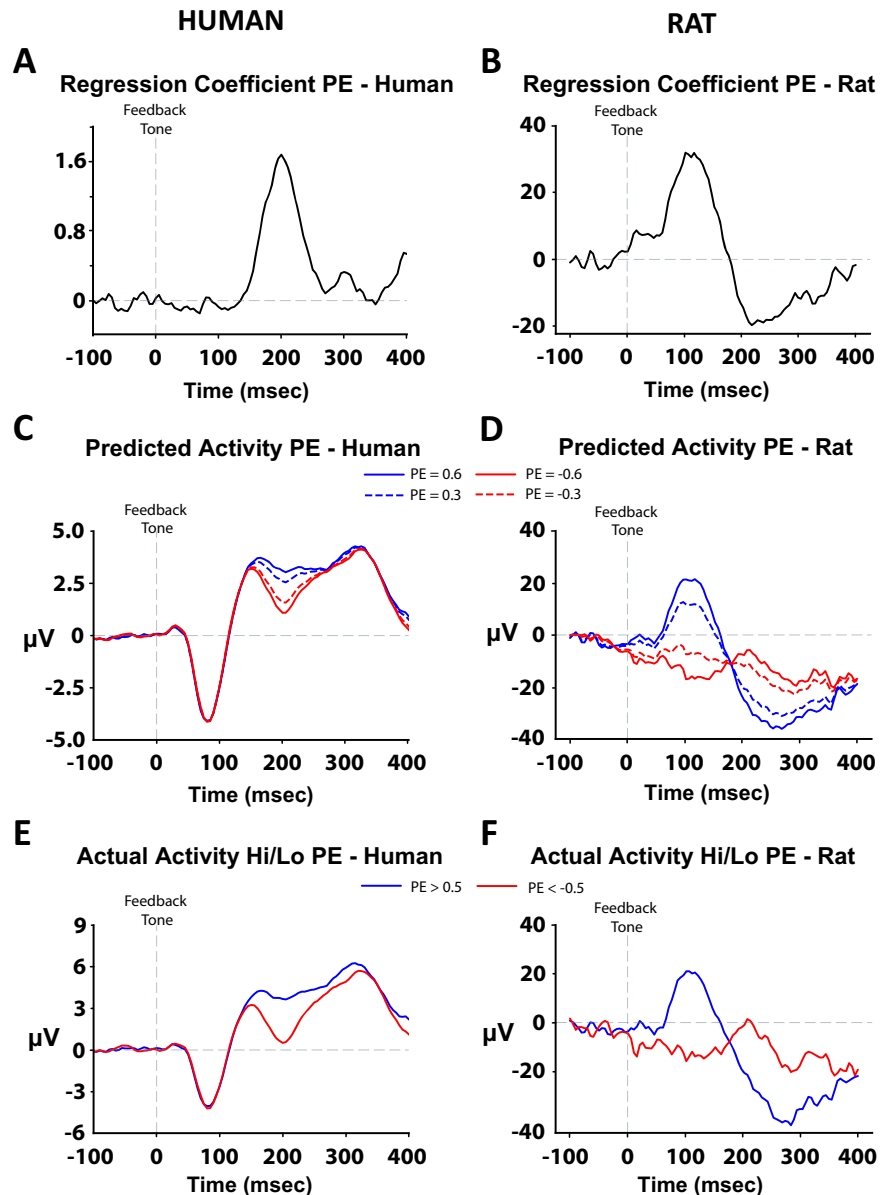
### Modafinil effects on frontal electrophysiological signals and PE values

To understand whether electrophysiological markers of cognitive flexibility are modulated by altering dopamine transmission, we administered the dopamine transport blocker modafinil to a separate cohort of humans ( $N = 29$ ) and the same cohort of rats described above ( $N = 11$ ) prior to PRL testing and EEG recording. Under the placebo/vehicle condition, we successfully replicated the frontal RewP observed between unrewarded and rewarded responses in both humans (Fig. 4A) and rats (Fig. 4B). In humans, a 1-way ANOVA of the ERPs following target responses revealed a significant main effect of Feedback [ $F(1,26) = 10.16, p = 0.004$ ] (Fig. 4C). The same significant Feedback effect [ $F(1,9) = 15.08, p = 0.004$ ] was identified in the 1-way ANOVA of the rodent LFPs (Fig. 4D). In both species, ERPs/LFPs were more negative following unrewarded target responses relative to rewarded target responses, with, unexpectedly, no effect of modafinil (see Supplementary Fig. 6 for ERPs from all modafinil doses for both species).

We then fitted the same Q-learning algorithm to PRL performance. One-way ANOVA models examining PEs following target responses



**Fig. 3 | The strong relationship between ERP voltage following reward feedback and PE values was comparable between humans (left) and rats (right).** Linear regression revealed a strong positive correlation between ERP voltage during the reward positivity (RewP) period and PE value in both humans (A) and rats (B). Data reflect average regression coefficient values. We then predicted the ERP response based on several fixed PE values (PE = 0.3, blue dashed line; PE = 0.6, blue solid line; PE = -0.3, red dashed line; PE = -0.6, red solid line) and showed that the predicted ERP was more negative or positive when the PE value was negative or positive, respectively, in both humans (C) and rats (D). These effects only emerged during the RewP period and were nearly identical to the actual ERP voltages when split between high (PE > 0.5; blue solid line) and low PE (PE < -0.5; red solid line) values in both humans (E) and rats (F). Data reflect grand average ERP traces.



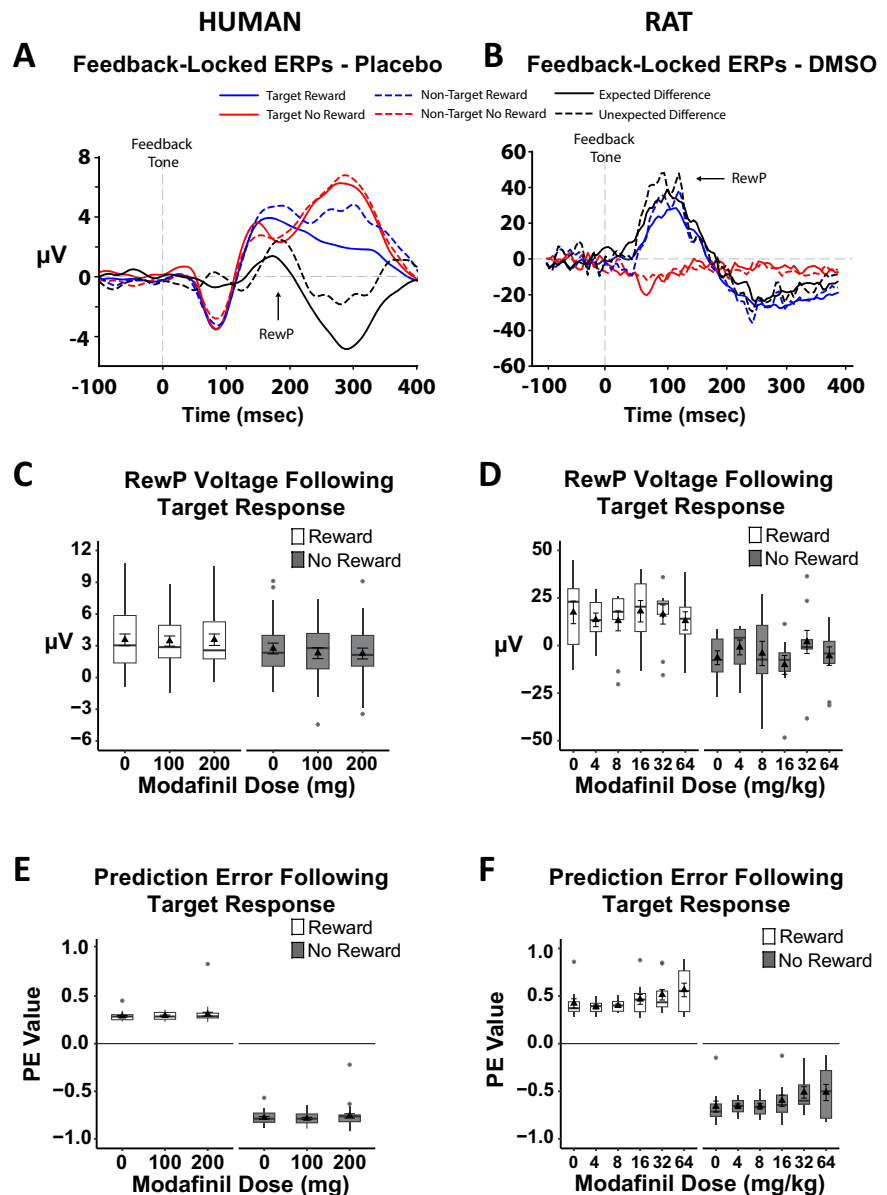
revealed a significant main effect of Feedback for both humans [ $F(1,28) = 7630.16$ ,  $p < 0.001$ ] (Fig. 4E) and rats [ $F(1,10) = 2737.01$ ,  $p < 0.001$ ] (Fig. 4F). Interestingly, in rats, a 1-way ANOVA across all modafinil doses revealed significant linear effects for the rewarded condition [ $F(5,50) = 2.62$ ,  $p = 0.035$ ] with higher doses resulting in more positive PEs; this was not the case for the unrewarded condition [ $F(5,50) = 1.90$ ,  $p = 0.111$ ]. Although the linear effects analysis of human PEs was not significant, the pattern of results was similar to rodents (i.e., increasing PEs with increasing modafinil dose). Although a wide dose response range was used in rodents, matching doses across species is challenging<sup>48</sup> and it is possible that the higher doses used in rats (e.g., 32 and 64 mg/kg) were relatively greater than the highest human dose. Thus, the pattern of increasing PEs across doses may have been more robust if the human participants were exposed to higher modafinil doses.

## Discussion

Using analogous PRL tasks across species, along with equivalent EEG data processing and identical computational analyses of behavioral and EEG data, we found remarkably concordant behavioral, computational, and electrophysiological findings across species, including a neurophysiological response consistent with a RewP in both humans and rats. In humans, the

RewP was observed over frontocentral brain regions, while the rodent RewP was recorded directly from the ACC. Notably, in humans, this component peaked approximately 200 ms following feedback, which is earlier than the more typically observed ~300 ms peak<sup>49</sup>. We postulate that this latency shift is likely due to using auditory feedback rather than the more typical visual feedback, but recognize that other task design effects may also be involved in this latency shift. Behavioral performance between species was comparable, as indicated by greater sensitivity to positive feedback (i.e., target win-stay) relative to negative feedback (i.e., target lose-shift), both of which significantly correlated with the overall number of reversals in both species. Although the number of reversals and win-stay responses were generally higher and lose-shift ratios were generally lower in humans compared to rats, these findings are consistent with previously published PRL data in both species<sup>40,50</sup>. Importantly, Q-learning computational analyses revealed that both humans and rats assigned greater value to target vs. non-target stimuli, and these values alternated between response apertures as the criterion for reversals was achieved and reward contingencies were switched. Consistent with our prior finding<sup>51</sup>, we found a positive correlation between the beta parameter and the number of completed reversals, likely because a higher beta parameter promotes a greater tendency to exploit the action with the higher value. Taken together, these findings confirm that humans and

**Fig. 4 | Modafinil did not alter feedback-locked ERPs, but did disrupt PE values in rats at high doses.** Testing with the placebo/vehicle produced a similar reward positivity (RewP) as previously shown (Fig. 2) in both humans (A) and rats (B). Modafinil did not alter the magnitude of the RewP in either species (C, D). However, positive PE values increased and negative PE values became less negative in response to high doses of modafinil in rats ( $n = 11$ ; F). A similar trend may have emerged in humans ( $n = 30$ ; E) at higher doses. Grand average ERP traces time-locked to the feedback tone are presented for all trial types (rewarded target = blue solid line; non-rewarded target = red solid line; rewarded non-target = blue dotted line; non-rewarded non-target = red dotted line) and expectancy-based difference waveforms (Expected = Rewarded Target – Non-rewarded Non-Target, and Unexpected = Rewarded Non-Target – Non-rewarded Target). Voltage and PE data are presented using box plots marking the median value (center line), the 25th and 75th percentiles (the outer edges), and  $\pm 1.5$  times the interquartile range (the whiskers), as well as the mean (triangle)  $\pm$  standard error of the mean (error bars).



rats performing the PRL task exhibit concordance in both behavioral performance and evoked neurophysiological responses.

Reward PEs are fundamental for learning<sup>52</sup>, and signal when an actual outcome deviates from what was expected. As the value assigned to the target was greater than the non-target, one would likely expect a reward after a target response. Importantly, rewarded outcomes elicited positive PEs, whereas non-rewards elicited negative PEs. However, the PE signal following a rewarded non-target response was more positive than rewarded target responses, and a non-rewarded target response elicited a more negative PE than non-rewarded non-target responses. Interestingly, the direction and magnitude of the frontal voltage across the RewP window following feedback were strikingly similar to the direction and magnitude of PEs across all four trial types in both species (Fig. 2C–F), and PEs significantly correlated with voltage during the RewP window in both species as well. These results suggest that the frontal voltage changes triggered by reward feedback represent a neural marker of PEs that is consistent across humans and rats. Less positive deflections after an unexpected reward omission are associated with greater negative PEs, whereas smaller deflections are associated with weaker negative PEs. Blunted negative PEs may contribute to perseverative responding and less flexibility if reward omissions fail to signal that an error has occurred and that alternate options

should be explored. Future research focused on understanding the trial-level relationship between neural activity and choice selection may provide further insight into the role of PEs in cognitive flexibility.

Because midbrain dopaminergic signals to the ACC may contribute to and/or modulate the RewP<sup>26,53</sup>, we hypothesized that indirectly enhancing dopamine levels (among other neuromodulators) via modafinil would enhance both the RewP and PEs. Interestingly, in rodents, we found that higher doses of modafinil increased positive PEs for rewarded target responses (i.e., PEs became more positive as modafinil dose increased) but suppressed negative PEs for unrewarded target responses (i.e., PEs became less negative as modafinil dose increased). The effect on PEs may disrupt the ability to properly identify and update the value of the target stimulus, ultimately impacting choice behavior. Notably, we only saw this effect in rodents administered the higher modafinil doses, and animals also completed fewer reversals and exhibited alterations in win-stay/lose-shift strategies (see Supplementary Table 2). The absence of this effect in humans suggests that the maximal human dose administered (200 mg) was not sufficient to disrupt behavior and neurophysiology in a manner consistent with that observed in rats.

It is possible that by increasing dopamine levels, high doses of modafinil might have interfered with the ability of midbrain dopamine neurons

to appropriately suppress activity during an unexpected event, thereby suppressing the PE, which is associated with an attenuated RewP. This is consistent with prior evidence indicating that increased dorsomedial striatum activity reduces the impact of loss, thereby impairing performance in a reversal learning task<sup>54</sup>. Critically, modafinil also modulates other monoamine systems<sup>55</sup>, such as noradrenaline and serotonin, which also play a role in regulating reward PEs and value updating<sup>56,57</sup>. Indeed, noradrenergic neurotransmission plays a central role in error monitoring<sup>58</sup>. Thus, it is unclear whether PRL disruptions evident in rats after the highest dose of modafinil are due to alterations in dopamine neurotransmission or another neurotransmitter system. Alternatively, another possible explanation for why modafinil did not modulate PEs and ERPs/LFPs may have to do with the subjects. All human participants were healthy and without any psychiatric conditions, and all rats were naïve and otherwise healthy. Thus, dopamine signaling in healthy subjects may already be optimal, and any additional increase may impede performance. It is possible that modafinil may improve PRL performance in subjects that are otherwise healthy but exhibit reduced task performance or in patient samples characterized by reduced cognitive flexibility. Future studies that are sufficiently powered could group participants based on task performance (i.e., optimal vs poor) before modafinil administration or consider neuropsychiatric samples.

Importantly, the results of the cross-species studies presented here highlight the potential to replicate the rodent study using a parallel patient sample. If successful, this and other similar cross-species approaches may be used to test novel putative pharmacotherapies using similar behavioral and neurophysiological measures with high predictive validity<sup>59</sup>. Such an approach would identify which neural mechanisms linked to behavioral endpoints are conserved across species and thus would be appropriate to assess using an animal model. Importantly, as a major step towards this goal, our findings demonstrate that concordant behavioral, computational, and neurophysiological measures are observed in humans and rodents performing a cross-species task. A similar approach could be used to identify neural mechanisms of and treatments for disorders featuring deficits in cognitive flexibility and other behaviors that can be accurately measured across species and linked to conserved neural processes.

## Methods

A detailed description of the multidisciplinary research program of which the present study was part is described at: [clinicaltrials.gov/study/NCT02855229](https://clinicaltrials.gov/study/NCT02855229).

### Humans subjects

Sixty-three volunteers, aged 18–45 years, were recruited for the first PRL cohort (which did not include modafinil testing). A total of 54 were retained (19 male, 35 female) for final behavioral data analyses, of which 34 (13 male, 21 female) were retained for EEG analyses after further exclusion due to poor EEG data quality. Thirty separate right-handed volunteers were recruited for the second PRL cohort (which included modafinil), and a total of 29 subjects (14 male, 15 female) were retained for final data analyses. Subjects were free of any psychiatric history, as determined by a clinician-administered Structured Clinical Interview for DSM-5 (SCID-5)<sup>60</sup>. Subjects were compensated \$452 for participation. All ethical regulations relevant to human research participants were followed, and all procedures were approved by the Mass General Brigham Institutional Review Board. Subjects provided written informed consent in the presence of a medical doctor prior to participation.

**Human PRL task procedure.** The first cohort participated in a single testing session. Subjects were randomly assigned to either the PRL task or a Flanker task (the results of which were reported separately)<sup>61</sup>. Analyses examining the effects of task order revealed no significant differences. The second PRL cohort completed three sessions, separated by at least one week, using a double-blind, within-subjects, placebo-controlled design; across sessions, subjects were administered 0 mg (placebo), 100 mg, or 200 mg modafinil (2 h pretreatment).

Subjects completed a modified version of the PRL task (Supplementary Fig. 1A) while seated ~70 cm from a computer monitor inside an acoustically and electrically shielded booth. All stimuli were presented on a 22.5-in VIEWPixx monitor (VPixx Technologies, Saint-Bruno, Canada) using PsychoPy software. In this PRL task, participants were tasked to choose between two stimuli, one of which had been randomly designated as the target stimulus at the beginning of the session. Participants received probabilistic feedback in that, if the target stimulus was chosen, a reward would follow 80% of the time. Similarly, if non-target stimuli, negative feedback was given 80% of the time. Thus, spurious feedback would occur 20% of the time. The target or non-target assignment was reversed if the participant selected the target stimulus on 8 consecutive trials irrespective of feedback.

To ensure parallel instructions between species, participants were not made aware of reversing contingencies. Rather, they were instructed that they would need to “choose between two circles in order to win as much money as you can” and that they would hear one tone indicating a win/correct selection or a different tone indicating an incorrect selection. Participants completed 10 practice trials to familiarize themselves with the trial structure and the two tones. One session consisted of 300 trials with one break after 150 trials.

Every trial started with the presentation of a fixation cross of random duration between 500 and 1000 ms. Stimuli consisted of a red and blue circle, randomly placed on the left or right side of the screen, presented for maximally 2000 ms or until a response was given. Participants selected the left circle by pressing “c” or the right circle by pressing “m” on a keyboard. As soon as the response was given, a black border appeared around the selected circle for 400 ms. Following a random delay of between 400 and 600 ms, auditory feedback (either a 700 Hz or 1000 Hz pure sine wave) was played for 200 ms. Assignment of the reward and omission outcome to the high or low tone was counterbalanced between participants. If the subject received positive feedback, the feedback tone was followed by the sound of a coin dropping for 1200 ms. This sound was added to mimic the consumption of the food reward in rodents. If participants failed to answer within 2000 ms, a 300 Hz tone was played together with a visual stimulus reading “No response!”.

**Human EEG acquisition.** In both PRL cohorts, EEG data were recorded using an actiCHamp amplifier and a 96 Ag/AgCl active electrode actiCAP system (Brain Products GmbH, Gilching, Germany) that used an equidistant spherical montage and was referenced online to a vertex channel (approximating Cz), with a ground electrode approximately at AFz. Data were digitized at 500 Hz using BrainVision Recorder, and impedances were kept below 35 kΩ.

### Rodent subjects

Eleven adult male ( $n = 5$ ) and female ( $n = 6$ ) Wistar rats were used for both baseline and modafinil experiments (Charles River Laboratories, Wilmington, MA, USA). Animals were pair-housed and food-restricted to 85% of their free-feeding body weight throughout behavioral training. After EEG electrodes were surgically implanted, animals were single-housed for the duration of the experiment and all PRL tests. All rats were housed in a vivarium room with a 12-h reverse light-dark cycle, with lights off between 7:00 AM and 7:00 PM. Rats were monitored daily for signs that would prompt a humane endpoint (e.g., excessive weight loss, inappetence, moribund state, or infection), requiring removal from the study and euthanasia. We have complied with all relevant ethical regulations for animal use; all rodent procedures were approved by the UC San Diego Institutional Animal Care and Use Committee, and were conducted in accordance with guidelines from the National Institute of Health and the Association for Assessment and Accreditation of Laboratory Animal Care.

**Rodent EEG surgery and data acquisition.** Prior to testing, rats were anesthetized with a 2% isoflurane/oxygen vapor mixture and secured on a stereotaxic frame (Kopf Instruments; Tujunga, CA, USA). In order to best approximate human EEG recordings, we implanted three different

electrode types over and within the rats' brains: (1) A 1/8" diameter fine silver disc (Hauser and Miller; St. Louis, MO, USA) soldered to a 0.01" diameter PFA-coated stainless steel wire (#792400; A-M Systems; Sequim, WA, USA) was placed on the surface of the skull immediately rostral to bregma; (2) a stainless steel jeweler's screw soldered to the wire described above was implanted in the skull over the frontal cortex (AP + 3.7 mm, ML ± 2.5 mm) and parietal cortex (AP −4.5 mm, ML ± 4.9 mm); (3) a stainless steel wire described above was inserted into the ACC (AP + 2.7 mm, ML ± 0.8 mm, DV −2.3, from bregma), lateral orbitofrontal cortex, nucleus accumbens shell, caudate nucleus, and primary auditory cortex; notably, electrode implantation angles were unified across animals. Recordings from sites other than the ACC will be reported separately. Reference and ground skull screws were implanted bilaterally over the cerebellum. All electrodes were secured initially with Denmat cement, then completely covered with dental acrylic. The wires from all electrodes were secured with gold pins into an electrode interface board (EIB-16; Neuralynx; Bozeman, MT, USA) that attached to a removable amplifier board during data acquisition. Rats were monitored for at least one-week post-surgery before EEG recording. Unifying electrode coordinates, materials, and implantation angles across animals was done to minimize variability in signal orientation<sup>62,63</sup>.

During testing, rats were connected to a 16-channel amplifier board (RHD2132; Intan Technologies; Los Angeles, CA, USA) that transmitted electrophysiological data to a USB interface board connected to a computer running RHD2000 interface software (Intan Technologies). Data were continuously recorded at a 1 kHz sampling rate and filtered between 0.1 and 300 Hz. While LFP data were being continuously collected during the PRL task, TTL event markers were recorded to identify presentation of reward feedback. Audio signals were recorded during testing and connected to the EEG acquisition system to confirm the accuracy of the time-lock between tones and neurophysiological signals.

**Rodent PRL task procedure.** Rats were trained and tested in a Plexiglas operant conditioning box (24 × 30 × 29 cm; Med Associates, St Albans, VT, USA) enclosed in a Faraday cage (Med Associates). It consisted of two retractable levers, a food receptacle positioned between the two levers, a stimulus light above each lever, a speaker above the food receptacle, and a house light placed 2 cm below the ceiling on the opposite wall. Tones were created by an audio generator. All programs and collection of data were done on MED-PC V software (Med Associates, St Albans, VT, USA).

The rodent PRL task was designed to be as similar to a human task as possible (Supplementary Fig. 1B). Briefly, rats responded for one of two colored light stimuli (that were illuminated for up to 5 s) by pressing one of two levers (presented 1 s after illumination) under the two lights. Target responses resulted in positive feedback (100 ms tone, 5 or 15 kHz, counterbalanced) on 80% of trials 500–1000 ms after the response and preceded the delivery of a 45 mg (for male rats) or 20 mg (for female rats) sucrose pellet (5TUT, Test Diet), or negative feedback on 20% of trials (other tone) followed by no pellet. Non-target responses resulted in negative and positive feedback on 80% and 20% of trials, respectively. Eight consecutive target responses, regardless of feedback, resulted in the target stimulus switching to the other light. A "reversal" was recorded when a rat successfully made 8 consecutive target responses after a switch. Rats completed 300 trials per session.

EEG recordings were obtained on the 21st day of testing (i.e., after rats had sufficient time to learn the PRL procedure). After one week, EEG recordings were obtained during testing after administration of one of the following doses of modafinil using a within-subjects Latin-square design: 0, 4, 8, 16, 32, 64 mg/kg. The vehicle for modafinil dosing was DMSO, administered at a volume of 1 ml/kg. There was a minimum one-week washout period between tests.

### Cross-species Task Performance and ERP/LFP derivation

In both species, in addition to reversals, target win-stay probabilities were calculated as the number of responses repeated after a rewarded target

response divided by the total number of rewarded target responses. Target lose-shift probabilities were calculated as the number of responses not repeated after an unrewarded target response divided by the total number of unrewarded target responses. Q-learning parameters, including Q, PE, alpha, beta, and forget values, were also calculated identically in both species (see Supplementary Methods).

All EEG data were analyzed with BrainVision Analyzer 2.1 (Brain Products GmbH, Gilching, Germany) in the following steps: human data were visually inspected to identify gross muscle activity and artifactual channels, and rodent data were checked for polarity inversions. Following this, data were bandpass filtered from 0.1 (12 dB/oct) to 30 Hz (24 dB/Oct Human cohort 1, 48 dB/oct, Human cohort 2, and Rats) using zero-phase Butterworth IIR filters. Human data were then subjected to independent component analysis to remove eye movement and EKG sources, spherical spline interpolation to replace corrupted channels, and finally re-referenced to the common average. Rodent data were re-referenced to the electrode implanted above the left cerebellum. Processed data for both species were then segmented into −1500 to 2000 ms epochs around the feedback stimulus, and segments were rejected channel-wise as artifact if any of these criteria were met: (1) a voltage exceeding ±75 μV (humans) or ±800 μV (rats); (2) a maximum voltage difference of less than 0.5 μV for more than 100 ms within a trial. Human recordings were also checked against two extra criteria: (1) a voltage step exceeding 50 μV, and (2) a maximum voltage difference of more than 150 μV across 200 ms time intervals within a trial.

In both species, feedback-locked data were segmented into individual epochs spanning from 200 ms before and 600 ms after the feedback tone, baseline-corrected (described below), and averaged. In humans, feedback-locked ERPs were quantified at channel 2 (approximating electrode FCz), and baseline-corrected to the −200 to 0 ms pre-feedback window. The RewP was quantified as the average amplitude between 165 and 225 ms post-feedback. In rats, feedback-locked LFPs were baseline-corrected to the −300 to −100 ms pre-feedback window and quantified at the ACC LFP channel as the average activity across the 60–160 ms post-feedback window.

### Statistics and reproducibility

Across all cohorts (i.e., 1 rodent ( $n = 11$ ) and 2 humans ( $n = 34$  and  $n = 30$ )), the associations between task performance and behavioral parameters were assessed using Pearson's correlation. In turn, between-condition comparisons for all parameters, at both the trial- and session-level, were evaluated using 1- and 2-way ANOVAs. Finally, the association between the ERP/LFP voltage and PE values was evaluated using a series of generalized linear models (GLMs) with a Gaussian distribution and identity link function. The general structure of these models was as follows:

$$E(\text{Voltage}_{dt}) = \beta_0 + \beta_1 PE_t \quad (1)$$

where  $E(\text{Voltage}_{dt})$  corresponds to the ERP/LFP amplitude at each data point,  $d$ , within a given trial,  $t$ , and  $PE_t$  is the signed PE value on that given trial.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Numerical source data used in the present manuscript are hosted on OSF and available at: [osf.io/gxkdv](https://osf.io/gxkdv). All other data (i.e., raw EEG recordings) that support the findings of this study are available from the corresponding author upon reasonable request.

### Code availability

The following software were used in preprocessing raw data and subsequent analysis: BrainVision Analyzer (v2.0), Python (v3.7.1), NumPy Python library (v1.21.5), pandas Python Library (v1.1.5), SciPy Python Library (v1.4.1), matplotlib Python library (v3.5.3), statsmodels Python Library



(v0.13.5), IBM SPSS Statistics (v24), and GraphPad Prism (v8). Where possible, Python and R code used in the present analysis are available on OSF at: [osf.io/gxkdv](https://osf.io/gxkdv). EEG preprocessing templates are available from the corresponding author on reasonable request.

Received: 18 June 2024; Accepted: 14 August 2025;  
Published online: 23 August 2025

## References

1. Scott, W. A. Cognitive complexity and cognitive flexibility. *Sociometry* **25**, 405 (1962).
2. Uddin, L. Q. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nat. Rev. Neurosci.* **22**, 167–179 (2021).
3. Zühlsdorff, K., Dalley, J. W., Robbins, T. W. & Morein-Zamir, S. Cognitive flexibility: neurobehavioral correlates of changing one's mind. *Cereb. Cortex* **33**, 5436–5446 (2023).
4. Genet, J. J. & Siemer, M. Flexible control in processing affective and non-affective material predicts individual differences in trait resilience. *Cogn. Emot.* **25**, 380–388 (2011).
5. Chen, Q. et al. Association of creative achievement with cognitive flexibility by a combined voxel-based morphometry and resting-state functional connectivity study. *Neuroimage* **102**, 474–483 (2014).
6. Cools, R. Neuropsychopharmacology of Cognitive Flexibility. In *Brain Mapping* 349–353 (Elsevier, 2015).
7. Gabrys, R. L., Tabri, N., Anisman, H. & Matheson, K. Cognitive control and flexibility in the context of stress and depressive symptoms: the cognitive control and flexibility questionnaire. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2018.02219> (2018).
8. Davis, J. C., Marra, C. A., Najafzadeh, M. & Liu-Ambrose, T. The independent contribution of executive functions to health related quality of life in older women. *BMC Geriatr.* **10**, 16 (2010).
9. Robinson, O. J., Cools, R., Carlisi, C. O., Sahakian, B. J. & Drevets, W. C. Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. *Am. J. Psychiatry* **169**, 152–159 (2012).
10. Taylor Tavares, J. V. et al. Neural basis of abnormal response to negative feedback in unmedicated mood disorders. *Neuroimage* **42**, 1118–1126 (2008).
11. Veiel, H. O. F. A preliminary profile of neuropsychological deficits associated with major depression. *J. Clin. Exp. Neuropsychol.* **19**, 587–603 (1997).
12. Reddy, L. F., Waltz, J. A., Green, M. F., Wynn, J. K. & Horan, W. P. Probabilistic Reversal Learning in Schizophrenia: Stability of Deficits and Potential Causal Mechanisms. *Schizophr. Bull.* **42**, 942–951 (2016).
13. Waltz, J. A. & Gold, J. M. Probabilistic reversal learning impairments in schizophrenia: further evidence of orbitofrontal dysfunction. *Schizophr. Res.* **93**, 296–303 (2007).
14. Chamberlain, S. R., Fineberg, N. A., Blackwell, A. D., Robbins, T. W. & Sahakian, B. J. Motor inhibition and cognitive flexibility in obsessive-compulsive disorder and trichotillomania. *Am. J. Psychiatry* **163**, 1282–1284 (2006).
15. Remijne, P. L. et al. Reduced orbitofrontal-striatal activity on a reversal learning task in obsessive-compulsive disorder. *Arch. Gen. Psychiatry* **63**, 1225 (2006).
16. Camchong, J. et al. Frontal hyperconnectivity related to discounting and reversal learning in cocaine subjects. *Biol. Psychiatry* **69**, 1117–1123 (2011).
17. Ragland, J. D. et al. CNTRICS final task selection: long-term memory. *Schizophr. Bull.* **35**, 197–212 (2009).
18. Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H. & Holmes, A. The neural basis of reversal learning: an updated perspective. *Neuroscience* **345**, 12–26 (2017).
19. Robbins, T. W. Cross-species studies of cognition relevant to drug discovery: a translational approach. *Br. J. Pharmacol.* **174**, 3191–3199 (2017).
20. Murphy, F. C., Michael, A., Robbins, T. W. & Sahakian, B. J. Neuropsychological impairment in patients with major depressive disorder: the effects of feedback on task performance. *Psychol. Med.* **33**, 455–467 (2003).
21. Chamberlain, S. R. et al. Orbitofrontal dysfunction in patients with obsessive-compulsive disorder and their unaffected relatives. *Science* (1979). **321**, 421–422 (2008).
22. Kehagia, A. A., Murray, G. K. & Robbins, T. W. Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. *Curr. Opin. Neurobiol.* **20**, 199–204 (2010).
23. Proudfit, G. H. The reward positivity: from basic research on reward to a biomarker for depression. *Psychophysiology* **52**, 449–459 (2015).
24. Kehler, P., Brigman, J. L. & Cavanagh, J. F. Depth recordings of the mouse homologue of the reward positivity. *Cogn. Affect Behav. Neurosci.* **24**, 292–301 (2024).
25. Miltner, W. H. R., Braun, C. H. & Coles, M. G. H. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a “generic” neural system for error detection. *J. Cogn. Neurosci.* **9**, 788–798 (1997).
26. Holroyd, C. B. & Coles, M. G. H. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* **109**, 679–709 (2002).
27. Yasuda, A., Sato, A., Miyawaki, K., Kumano, H. & Kuboki, T. Error-related negativity reflects detection of negative reward prediction error. *Neuroreport* **15**, 2561–2565 (2004).
28. Chase, H. W., Swainson, R., Durham, L., Benham, L. & Cools, R. Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *J. Cogn. Neurosci.* **23**, 936–946 (2011).
29. Holroyd, C. B., Pakzad-Vaezi, K. L. & Krigolson, O. E. The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology* **45**, 688–697 (2008).
30. Holroyd, C. B., Krigolson, O. E. & Lee, S. Reward positivity elicited by predictive cues. *Neuroreport* **22**, 249–252 (2011).
31. Hauser, T. U. et al. The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage* **84**, 159–168 (2014).
32. Gruendler, T. O. J., Ullsperger, M. & Huster, R. J. Event-related potential correlates of performance-monitoring in a lateralized time-estimation task. *PLoS ONE* **6**, e25591 (2011).
33. Hyman, J. M., Holroyd, C. B. & Seamans, J. K. A novel neural prediction error found in anterior cingulate cortex ensembles. *Neuron* **95**, 447–456.e3 (2017).
34. Montague, P., Dayan, P. & Sejnowski, T. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
35. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
36. Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
37. Brown, J. W. & Braver, T. S. Learned predictions of error likelihood in the anterior cingulate cortex. *Science* **307**, 1118–1121 (2005).
38. Cohen, M. X. & Ranganath, C. Reinforcement learning signals predict future decisions. *J. Neurosci.* **27**, 371–378 (2007).
39. Cavanagh, J. F. et al. Electrophysiological biomarkers of behavioral dimensions from cross-species paradigms. *Transl. Psychiatry* **11**, 482 (2021).

40. Bari, A. et al. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology* **35**, 1290–1301 (2010).
41. Tranter, M. M., Aggarwal, S., Young, J. W., Dillon, D. G. & Barnes, S. A. Reinforcement learning deficits exhibited by postnatal PCP-treated rats enable deep neural network classification. *Neuropsychopharmacology* **48**, 1377–1385 (2023).
42. Cavanagh, J. F. et al. Amphetamine alters an EEG marker of reward processing in humans and mice. *Psychopharmacology* **239**, 923–933 (2022).
43. Volkow, N. D. et al. Effects of modafinil on dopamine and dopamine transporters in the male human brain. *JAMA* **301**, 1148 (2009).
44. Tranter, M. M. et al. Postnatal phencyclidine-induced deficits in decision making are ameliorated by optogenetic inhibition of ventromedial orbitofrontal cortical glutamate neurons. *Biol. Psychiatry Glob. Open Sci.* **4**, 264–274 (2024).
45. Cohen, J. D., McClure, S. M. & Yu, A. J. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 933–942 (2007).
46. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* (1979). **275**, 1593–1599 (1997).
47. Daw, N. D. & Doya, K. The computational neurobiology of learning and reward. *Curr. Opin. Neurobiol.* **16**, 199–204 (2006).
48. Sharma, V. & McNeill, J. H. To scale or not to scale: the principles of dose extrapolation. *Br. J. Pharmacol.* **157**, 907–921 (2009).
49. Sambrook, T. D. & Goslin, J. A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychol. Bull.* **141**, 213–235 (2015).
50. Rychlik, M., Bollen, E. & Rygula, R. Ketamine decreases sensitivity of male rats to misleading negative feedback in a probabilistic reversal-learning task. *Psychopharmacology* **234**, 613–620 (2017).
51. Barnes, S. A. et al. Modulation of ventromedial orbitofrontal cortical glutamatergic activity affects the explore-exploit balance and influences value-based decision-making. *Cereb. Cortex* **33**, 5783–5796 (2023).
52. Millard, S. J., Bearden, C. E., Karlsgodt, K. H. & Sharpe, M. J. The prediction-error hypothesis of schizophrenia: new data point to circuit-specific changes in dopamine activity. *Neuropsychopharmacology* **47**, 628–640 (2022).
53. Warren, C. M., Hyman, J. M., Seamans, J. K. & Holroyd, C. B. Feedback-related negativity observed in rodent anterior cingulate cortex. *J. Physiol. Paris* **109**, 87–94 (2015).
54. Young, M. K. et al. Activity in the dorsomedial striatum underlies serial reversal learning performance under probabilistic uncertainty. *Biol. Psychiatry Glob. Open Sci.* **3**, 1030–1041 (2023).
55. Minzenberg, M. J. & Carter, C. S. Modafinil: a review of neurochemical actions and effects on cognition. *Neuropsychopharmacology* **33**, 1477–1502 (2008).
56. Basu, A. et al. Frontal norepinephrine represents a threat prediction error under uncertainty. *Biol. Psychiatry* **96**, 256–267 (2024).
57. Luo, Q. et al. Comparable roles for serotonin in rats and humans for computations underlying flexible decision-making. *Neuropsychopharmacology* **49**, 600–608 (2024).
58. Riba, J., Rodríguez-Fornells, A., Morte, A., Münte, T. F. & Barbanoj, M. J. Noradrenergic stimulation enhances human action monitoring. *J. Neurosci.* **25**, 4370–4374 (2005).
59. Der-Avakian, A., Barnes, S. A., Markou, A. & Pizzagalli, D. A. Translational assessment of reward and motivational deficits in psychiatric disorders. *Curr. Top. Behav. Neurosci.* **28**, 231–262 (2015).
60. First, M. B., Williams, J. B. W., Karg, R. S. & Spitzer, R. L. Structured clinical interview for DSM-5 — research version (SCID-5 for DSM-5, Research Version; SCID-5-RV). (Arlington, VA, American Psychiatric Association, 2015).
61. Robble, M. A. et al. Concordant neurophysiological signatures of cognitive control in humans and rats. *Neuropsychopharmacology* **46**, 1252–1262 (2021).
62. Einevoll, G. T., Kayser, C., Logothetis, N. K. & Panzeri, S. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat. Rev. Neurosci.* **14**, 770–785 (2013).
63. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012).

## Acknowledgements

D.A.P. discloses support for the research and publication of this work from the National Institute of Mental Health grants UH2 MH109334 and UH3 MH109334. The authors would like to acknowledge the members of our scientific advisory board, Dr. Cindy Ehlers, Dr. Stan Floresco, Dr. Patricio O'Donnell, and Dr. Steven Siegel, for their assistance in the development and execution of these studies. We also thank Ms. Jessica Dally for technical assistance.

## Author contributions

A.D., J.B., W.A.C., V.B.R., S.L., D.A.P. contributed to the study design, A.D., H.S.S., B.D.K., S.L., S.N., M.A.R., M.B., A.M.I., R.L., S.P., E.F.C., G.P.N., J.P.K. conducted the experiments, S.A.B., Z.W., H.P., D.G.D., E.M. performed data analyses, A.D., S.A.B., T.L., H.S.S. wrote the manuscript.

## Competing interests

Over the past 3 years, B.D.K. has had sponsored research agreements with BlackThorn Therapeutics, Compass Pathways, Delix Therapeutics, Engrail Therapeutics, Neurocrine Biosciences, and Takeda Pharmaceuticals. Over the past 3 years, V.B.R. has received consulting fees from Engrail Pharmaceuticals, Jazz Pharmaceuticals, and Cohen Veterans Biosciences. Over the past 3 years, D.A.P. has received consulting fees from Arrowhead Pharmaceuticals, Boehringer Ingelheim, Compass Pathways, Engrail Therapeutics, Neumora Therapeutics, Neurocrine Biosciences, Neuroscience Software, and Takeda; he has received honoraria from the American Psychological Society, Psychonomic Society and Springer (for editorial work) and from Alkermes; he has received research funding from the Bird Foundation, Brain and Behavior Research Foundation, Dana Foundation, DARPA, Millennium Pharmaceuticals, the National Institute of Mental Health, and Wellcome Leap; he has received stock options from Ceretype Neuromedicine, Compass Pathways, Engrail Therapeutics, Neumora Therapeutics, and Neuroscience Software. No funding or any involvement from these entities was used to support the current work, and all views expressed are solely those of the authors. All other authors have no conflicts of interest or relevant disclosures.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-08729-x>.

**Correspondence** and requests for materials should be addressed to Diego A. Pizzagalli.

**Peer review information** *Communications Biology* thanks Clay B Holroyd, James Cavanagh, and Markus Ullsperger for their contribution to the peer review of this work. Primary Handling Editors: Christian Beste and Benjamin Bessieres. [A peer review file is available.]

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

## Supplemental Materials

### Supplementary Methods

#### *Q-learning modeling of PRL behavioral data and quantification of model performance*

We fit a variety of Q-learning models to the behavioral data. On each trial ( $t$ ), the value ( $Q$ ) of the chosen action ( $c$ ) is estimated according to the outcome delivered ( $r$ ; reward = 1, no reward = 0).  $Q$  values were initialized to 0.5 (neither good nor bad since the target and non-target stimuli were randomly assigned at the beginning of the session). Prediction errors (PE) — the difference between the estimated value of the chosen action ( $Q_c$ ) and the actual outcome — were computed according to:

$$PE = r_{(t)} - Q_{c(t)} \quad (1)$$

The PE was then used to update the value estimate for the chosen action (but not the non-chosen action) on a trial-by-trial basis according to equation 2.

$$Q_{c(t+1)} = Q_c + \alpha \times PE \quad (2)$$

The  $\alpha$  parameter is a learning rate, ranging from 0 to 1, that controls how quickly the PE updates value estimates. A low  $\alpha$  results in  $Q$ -values that gradually changed over multiple trials, whereas a high  $\alpha$  causes more rapid changes in  $Q$ -values.

The above equations are used to modify the value attributed to each action. The values are used to guide choices by converting them into action probabilities using the softmax equation.

$$p(A) = \frac{e^{QA \times \beta}}{e^{QA \times \beta} + e^{QB \times \beta}} \quad (3)$$

The degree to which choices are exploratory (i.e., selecting the lower-valued action) vs. exploitative (i.e., selecting the higher-valued action) is controlled by the inverse temperature ( $\beta$ ) parameter. A higher  $\beta$  parameter indicates a greater tendency to engage in exploitative choices



whereas a lower  $\beta$  parameter is indicative of a greater tendency to explore actions associated with a lower value. The model above describes the simplest model that we evaluated and includes two free parameters ( $\alpha$ ,  $\beta$ ) that are estimated.

We implemented several modifications to determine if any model variant provided a better explanation of behavior. As one example, PRL performance might involve simultaneously updating the value of both the chosen and the non-chosen actions after each PE. To evaluate this possibility, we implemented a model with a “double update” rule, whereby the value of both the chosen and unchosen actions are updated by the PE<sup>1</sup>. The value of the chosen action is updated according to equation 2 but the value of the unchosen action ( $Q_{nc}$ ) is updated according to:

$$Q_{nc(t+1)} = Q_{nc} - \alpha \times PE \quad (4)$$

With this variant, a rewarded left response would increase the value attributed to the left action while simultaneously decreasing the value associated with the right action. Alternatively, another approach that modulates the value of both actions on the same trial incorporates a method to decay the value of the unchosen action. This approach is designed to capture “forgetting” the value of an action if it has not been selected for several trials<sup>2,3</sup>. Again, the value of the chosen action is updated according to equation 2, but the value of the unchosen action is updated according to:

$$Q_{nc(t+1)} = (1 - forget) \times Q_{nc} \quad (5)$$

When the forget parameter equals zero, then the value of the unchosen action is unaffected, and the model is identical to that described by equation 2. By contrast, if the subject chooses the left action and the forget parameter equals 0.5, then the value of the right action is reduced by 50% for each trial where the left action was selected.

Each of the model variants above were used in conjunction with the softmax function (described by equation 3) to convert value estimates into choice probabilities. Although the

inverse temperature ( $\beta$ ) parameter regulates the degree to which the subject engages in explorative vs. exploitative decisions, choice behavior could also be influenced by the presence of a bias for one action or another. Therefore, rather than using the standard softmax function (equation 3), each of the model variants described above can also be used in conjunction with a modified softmax function that includes a bias parameter:

$$p(A) = \frac{e^{(QA+bias) \times \beta}}{e^{(QA+bias) \times \beta} + e^{QB \times \beta}} \quad (6)$$

A bias parameter of 1 would indicate an extreme bias for action A and a bias parameter of -1 would indicate an extreme bias for action B. When the bias parameter was 0 then the subject exhibited no bias for either action and equation 6 would function in a similar manner to the softmax function described in equation 3.

The models described above all used a single  $\alpha$  parameter to control the rate at which PEs updated value estimates. To explore whether a difference in the sensitivity for rewards vs. non-rewards influenced PRL performance, we also implemented a series of models that used separate learning rates for each outcome, as follows:

$$Qc_{(t+1)} = Qc + \alpha - gain \times PE \text{ if reward} = 1 \quad (7)$$

$$Qc_{(t+1)} = Qc + \alpha - loss \times PE \text{ if reward} = 0 \quad (8)$$

If the response resulted in a reward, then the rate at which PEs updated value estimates was controlled by the positive learning rate ( $\alpha$ -gain; equation 7). By contrast, for non-rewarded trials a negative learning rate was used ( $\alpha$ -loss; equation 8).

Using the equations described above, we fit 10 separate models to the PRL data. The optimal parameter values for each of the models were identified by minimizing the negative log-likelihood of choice probabilities using the “minimize” optimization function with the L-BFGS-B algorithm in Python’s Scipy library (v.1.5)<sup>4</sup>. To sample broadly from the range of possible

parameter values and reduce the possibility of the model getting stuck in local minima, the optimization algorithm was initialized at five different starting points within the parameter space for each subject. The parameter values associated with the lowest negative log-likelihood were selected as the best-fitting set for each subject.

The best-fitting model was determined by comparing the Bayesian Information Criteria (BIC) value for each model. The  $\Delta BIC$  value was calculated using the difference between the BIC of each model from the lowest BIC value and used to calculate the relative likelihood as follows:

$$relative\ likelihood = e^{(-0.5 \times \Delta BIC)} \quad (9)$$

To confirm the validity of the best-fitting model, we conducted posterior predictive checks in which we compared the simulated performance with actual PRL performance for human and rodent subjects. Parameter recovery exercises were also performed to confirm the accuracy of parameter estimation<sup>5</sup>.

To determine the relationship between the ERPs and reward PEs, we used generalized linear models to predict the trial-by-trial neural activity at each timepoint within a trial. The first regression model presented in Figure 3 included the reward PE for each trial as the sole model predictor whereas the regression model presented in Supplemental Figure 5 included the outcome of the trial (i.e., rewarded or not) and the value of the chosen action (i.e., the expected value). This approach is designed to delineate between error signaling and outcome valence<sup>6,7</sup>. In both models, neural activity for each timepoint within the trial epoch was used as the dependent variable. To predict the ERP waveform for a given reward PE value, we multiplied a hypothetical reward PE value (e.g., -0.6, -0.3, 0.3, or 0.6) with the regression coefficient and added the model intercept. This step was repeated for each timepoint.

## Supplementary Tables

### Supplementary Table 1. Quantification of Model Performance

Model performance was quantified and evaluated using BIC. The best-fitting model with the lowest BIC value for both species incorporated a single alpha learning rate and forget parameters (oaf). o=one; t=two; a=alpha; f=forget; du=double update; b=bias.

Model	Humans			Rats		
	BIC	$\Delta$ BIC	Relative Likelihood	BIC	$\Delta$ BIC	Relative Likelihood
oaf	239.963	0	1	341.830	0	1
tadu	240.117	0.154	0.92589	344.949	3.119	0.210241
oa	242.149	2.186	0.335209	345.204	3.374	0.185074
taf	243.634	3.671	0.159534	346.296	4.466	0.107206
oabf	244.164	4.201	0.122395	349.840	8.010	0.018224
oab	244.515	4.552	0.102694	351.546	9.716	0.007766
ta	244.852	4.889	0.08677	352.845	11.015	0.004056
oadu	246.246	6.283	0.043218	354.768	12.938	0.001551
tab	247.628	7.665	0.021655	355.848	14.018	0.000904
tabf	247.810	7.847	0.019772	357.235	15.405	0.000452



**Supplementary Table 2. Effect of modafinil on PRL performance**

In humans, modafinil treatment had no significant effect on the number of completed reversals, or win-stay or lose-shift responses for either response. By contrast, there was a significant dose on the number of completed reversals when rats were administered modafinil [ $F(5,45)=3.17$ ,  $p<0.05$ ], which was due to a reduction when administered 64 mg/kg ( $p<0.05$ ). Similarly, a dose effect was evident when target win-stay responding was analyzed [ $F(5,45)=3.05$ ,  $p<0.05$ ] that was also driven by a reduction in the highest treatment group ( $p<0.05$ ). Modafinil treatment had no significant effect on the remaining PRL measures.

Humans					
Dose	Completed Reversals	TWS	NTWS	TLS	NTLS
Placebo	4.01 $\pm$ 0.30	0.88 $\pm$ 0.01	0.82 $\pm$ 0.04	0.58 $\pm$ 0.04	0.67 $\pm$ 0.03
Low	4.40 $\pm$ 0.39	0.87 $\pm$ 0.01	0.84 $\pm$ 0.04	0.55 $\pm$ 0.04	0.67 $\pm$ 0.03
High	4.39 $\pm$ 0.30	0.88 $\pm$ 0.01	0.85 $\pm$ 0.03	0.52 $\pm$ 0.04	0.66 $\pm$ 0.03
Rats					
Vehicle	3.19 $\pm$ 0.28	0.78 $\pm$ 0.02	0.76 $\pm$ 0.04	0.44 $\pm$ 0.03	0.53 $\pm$ 0.03
4 mg/kg	3.35 $\pm$ 0.31	0.81 $\pm$ 0.01	0.80 $\pm$ 0.04	0.43 $\pm$ 0.02	0.51 $\pm$ 0.03
8 mg/kg	3.27 $\pm$ 0.39	0.78 $\pm$ 0.02	0.73 $\pm$ 0.03	0.47 $\pm$ 0.03	0.54 $\pm$ 0.02
16 mg/kg	2.64 $\pm$ 0.37	0.76 $\pm$ 0.03	0.79 $\pm$ 0.04	0.43 $\pm$ 0.06	0.51 $\pm$ 0.04
32 mg/kg	2.93 $\pm$ 0.41	0.73 $\pm$ 0.03	0.80 $\pm$ 0.03	0.49 $\pm$ 0.05	0.46 $\pm$ 0.04
64 mg/kg	1.76 $\pm$ 0.27	0.67 $\pm$ 0.05	0.76 $\pm$ 0.05	0.50 $\pm$ 0.06	0.46 $\pm$ 0.06

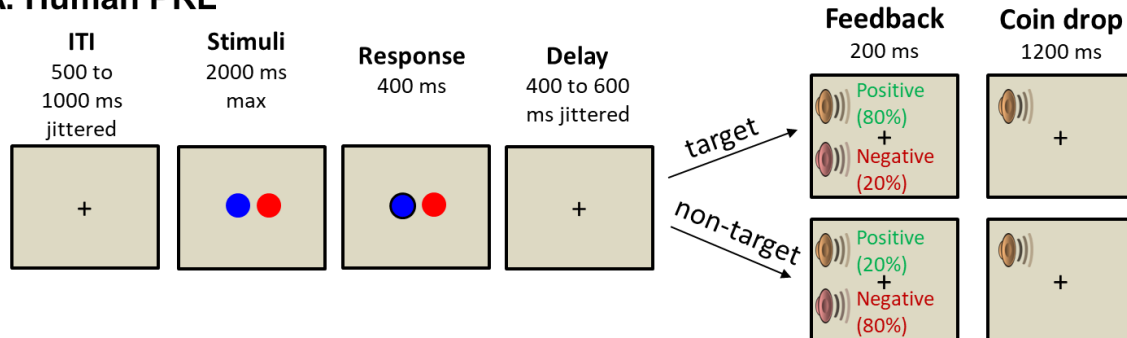
**Note:** NTLS = Non-Target Lose Shift; NTWS = Non-Target Win-Stay; TLS = Target Lose-Shift;  
TWS = Target Win-Stay

## Supplementary Figures

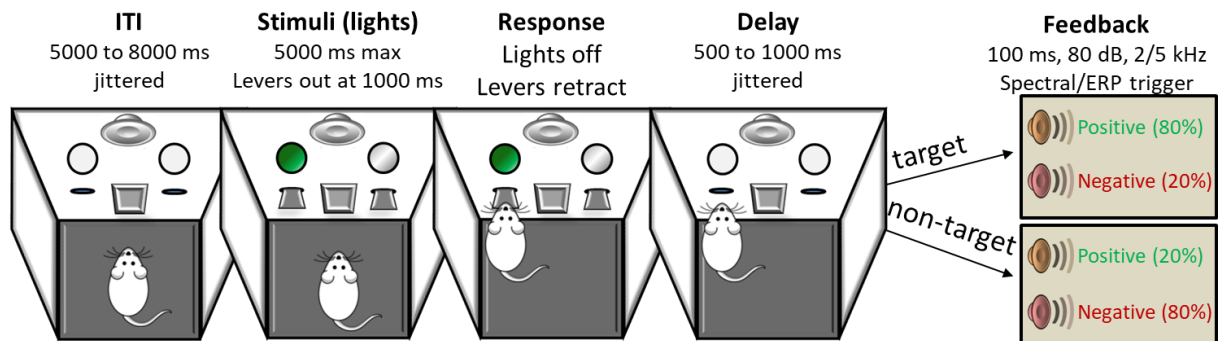
### Supplementary Figure 1. Cross-species PRL Task Design

The human (A) and rat (B) versions of the PRL task were designed to be procedurally similar. During each trial, two stimuli were presented simultaneously and subjects used a keyboard (humans) or levers (rats) to select one of the stimuli. One stimulus was randomly selected to be the target at the beginning of each test session. Selection of the target was rewarded with 80% probability, whereas selection of the non-target stimulus was rewarded with 20% probability. After each response, a high or low frequency tone signaled the subsequent presentation or omission of a reward (counterbalanced). When the target was selected on 8 consecutive trials, the reinforcement contingencies reversed such that the other stimulus became the target. Each test session lasted 300 trials for both species, and the total number of reversals and other measures (e.g., target win-stay and lose-shift probabilities) were recorded.

#### A. Human PRL

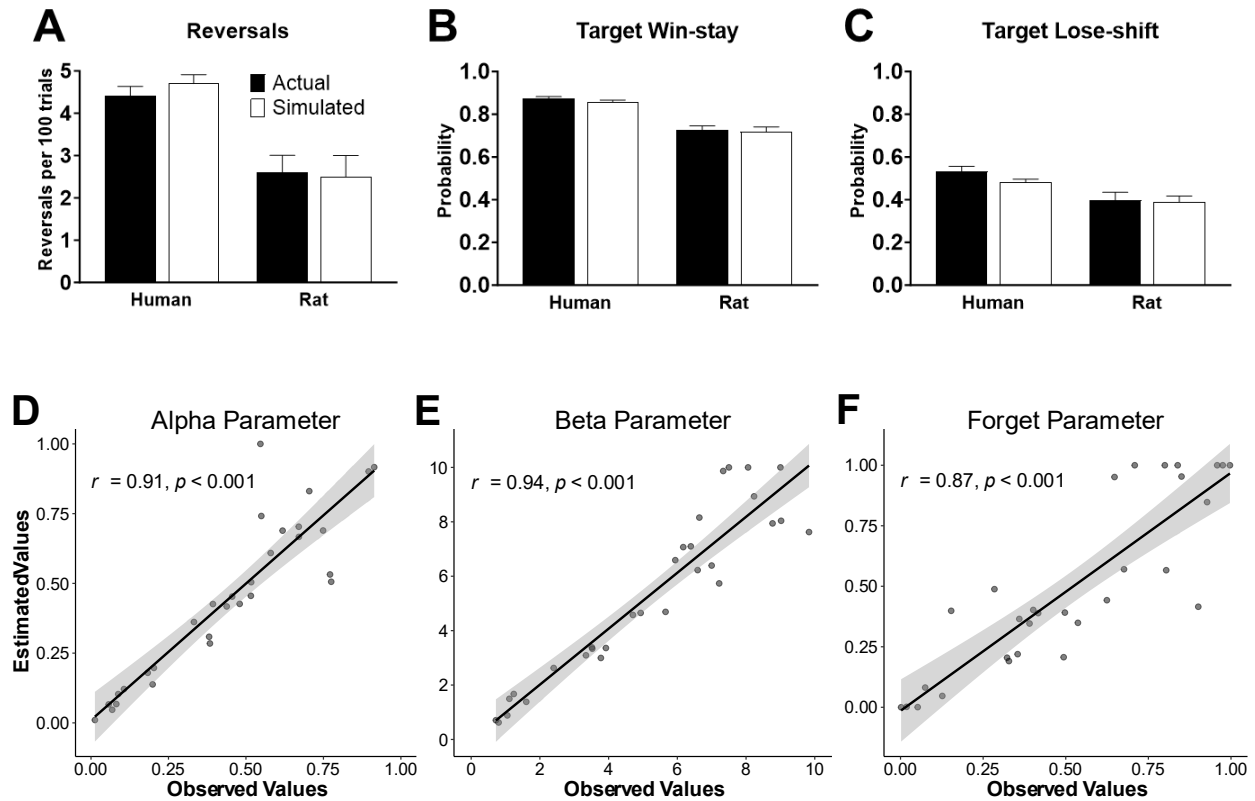


#### B. Rat PRL



### Supplementary Figure 2. Simulated vs. Actual PRL Behavioral Data

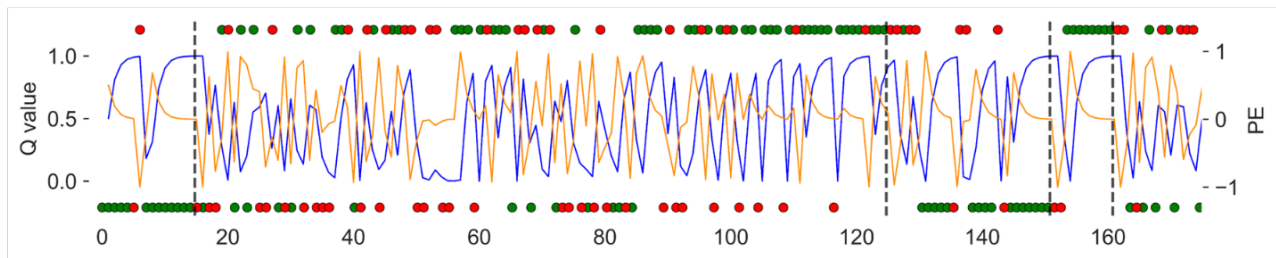
To confirm the validity of the selected Q-learning model we first performed a posterior predictive check by simulating PRL performance ( $n = 30$ ) using the parameter values obtained from fitting the model to behavior. There was a close correspondence between the actual and simulated PRL performance for the number of completed reversals (A), target win-stay (B), and target lose-shift responses (C). Additionally, we ensured that the three model parameters of the chosen model (alpha, beta, forget parameters) were accurately estimated. We simulated PRL performance using known values for each parameter and then fitted the Q-learning model to the simulated PRL data. A positive correlation for each of the three model parameters was evident between the simulated and fitted values (D – F), indicating that the parameters were recoverable and accurately estimated.



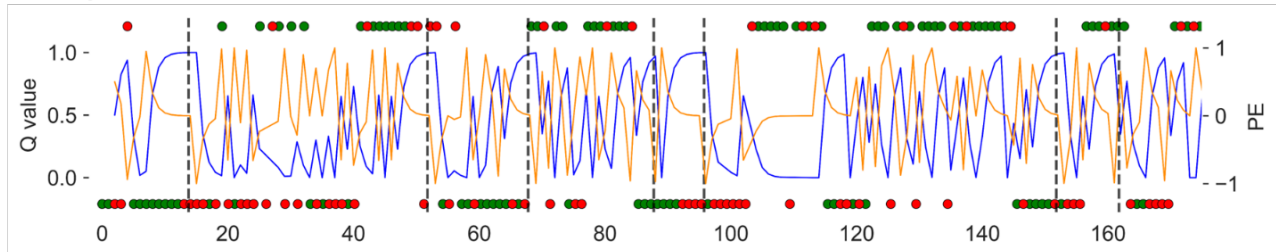
### Supplemental Figure 3. Q and PE Values During PRL Task Performance

Representative Q and PE values are displayed across a test session for humans (A) and rats (B). Blue and orange lines represent Q and PE values, respectively, across trials. The dots above and below Q and PE values represent responses on the left/right apertures. The dotted lines represent periods during which the response criterion is achieved and a reversal is recorded. Trial number is indicated along the x-axis. Across an entire test session, Q values were significantly greater for target vs. non-target stimuli in both humans ( $n = 54$ ) and rats ( $n = 11$ ) (C).

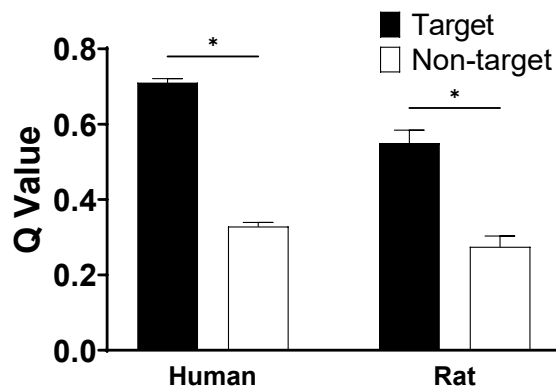
#### A. Representative Human Q and PE values



#### B. Representative Rat Q and PE values



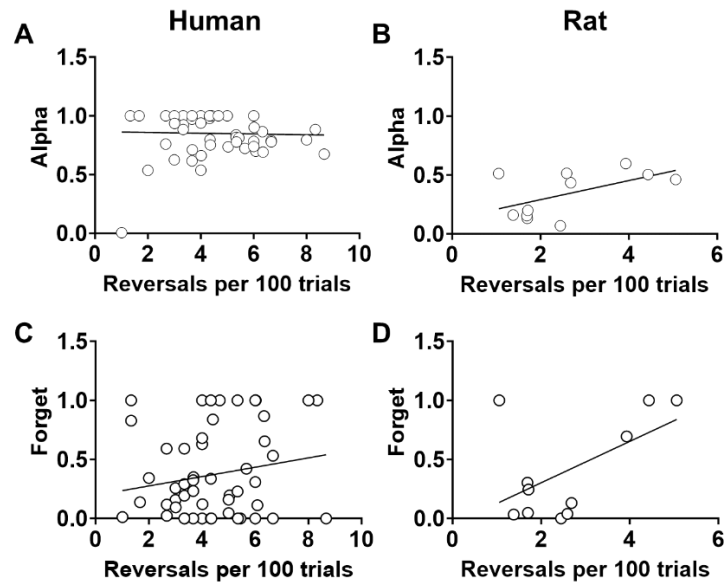
#### C. Average Q values for target and non-target stimuli





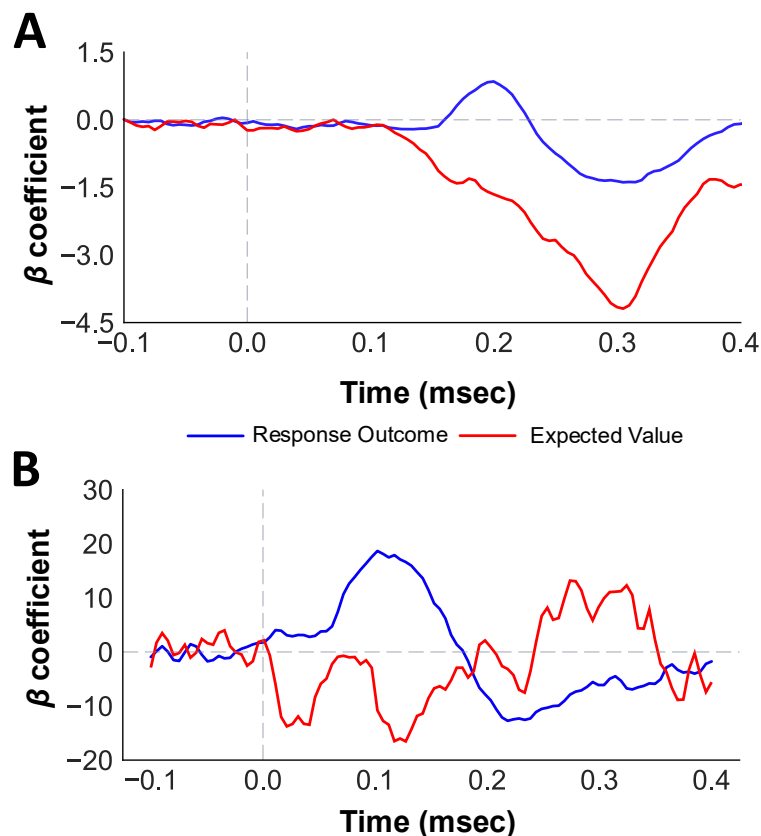
**Supplemental Figure 4. Q-learning Model Parameter Association with Reversals**

Alpha (A, B) and forget (C, D) parameters were correlated with number of reversals per 100 trials in both humans ( $n = 54$ ) and rats ( $n = 11$ ) using simple linear regression. None of these model parameters were significantly associated with reversals in either species (all  $r$  values  $< 0.55$ ; all  $p$  values  $> 0.05$ ).



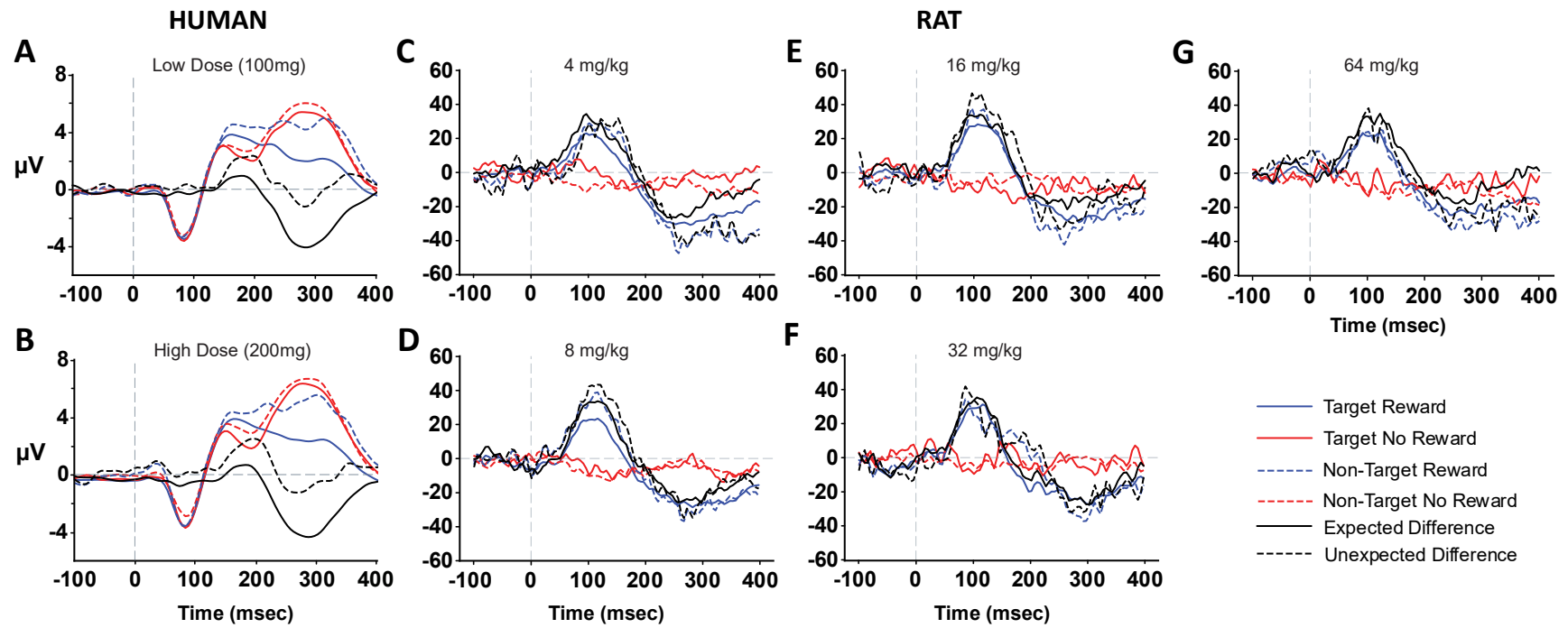
**Supplemental Figure 5. Relationship between neural activity and the outcome and expected value components of the PE**

The regression model presented in Figure 3 of the main text revealed a positive relationship between neural activity and reward PEs. However, the sign of reward PEs is dependent on the outcome received. Therefore, to confirm that this relationship was due to prediction error signaling and not simply due to the outcome valence, it was important to disentangle this potential collinearity by performing a secondary regression analysis that included outcome and expected value (i.e., the two sub-components of a reward prediction error) as model predictors. As expected, a positive regression coefficient was evident in humans (A) and rodents (B) for the response outcome (blue line) and a negative relationship was evident for expected value (red line). Importantly, this divergence coincided with the timepoint in the first model where reward PEs was positively associated with neural activity. Thus, this relationship is likely due to an association with prediction error signaling and not simply due to outcome valence.



**Supplemental Figure 6. Feedback locked ERP (human) and LFP (rodent) waveforms for each administered dose of modafinil**

Human ( $n = 30$ ) and rodent ( $n = 11$ ) subjects each completed the PRL task after being administered different doses of modafinil and EEG data were captured throughout these sessions. Human participants were administered 100 mg (A) and 200 mg (B) doses, while rodents were administered weight-based doses of 4mg (C), 8mg (D), 16 mg (E), 32 mg (F), and 64 mg (G) per kilogram of body weight. All plots present the evoked neural activity following the delivery of feedback for the various trial types of the task (rewarded target = blue solid line; nonrewarded target = red solid line; rewarded non-target = blue dotted line; non-rewarded non-target = red dotted line) and expectancy-based difference waveforms (Expected = Rewarded Target – Non-rewarded Non-Target, and Unexpected = Rewarded Non-Target – Non-rewarded Target) and identify a Reward Positivity (RewP) or RewP-like signal in humans and rats, respectively.



## Supplementary References

1. Schlagenhauf F. et al. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *NeuroImage* **89**, 171-80 (2014).
2. Katahira K. The relation between reinforcement learning parameters and the influence of reinforcement history on choice behavior. *Journal of mathematical psychology* **66**, 59-69 (2015).
3. Hattori R., Danskin B., Babic Z., Mlynaryk N. & Komiyama T. Area-specificity and plasticity of history-dependent value coding during learning. *Cell* **177(7)**, 1858-72.e15 (2019).
4. Weinstein A., El-Deredy W., Chabert S., Fuentes M. editors. Fitting human decision making models using python. *Proceedings of the 15th Python in Science Conference*. Ed. by Sebastian Benthall and Scott Rostrup, 1-6 (2016).
5. Wilson R.C. & Collins A.G.E. Ten simple rules for the computational modeling of behavioral data. *eLife* **8**:e49547 (2019).
6. Zhang L. & Gläscher J. A brain network supporting social influences in human decision-making. *Science Advances* **6**:eabb4159 (2020).
7. Barnes S.A., Dillon D.G., Young J.W., Thomas M.L., Faget L., Yoo J.H., Der-Avakian A., Hnasko T.S., Geyer M.A. & Ramanathan D.S. Modulation of ventromedial orbitofrontal cortical glutamatergic activity affects the explore-exploit balance and influences value-based decision-making. *Cerebral Cortex* **33(10)**, 5783-5796 (2023)