

Pupil-linked arousal encodes uncertainty-weighted prediction errors

Prashanti Ganesh^{1,2}, Tobias H. Donner^{3,4}, Radoslaw M. Cichy^{1,2}, Nicolas W. Schuck^{5,6},
Carsten Finke^{2,7}, & Rasmus Bruckner^{1,6,*}

Learning to predict future outcomes is essential for successful decision-making. One important mechanism governing such learning is the reward prediction error. In many real-world scenarios, sensory information about stimuli and choice options is ambiguous, leading to uncertainty about the environment's underlying states that guide learning and choice behavior. In such cases, learning from prediction errors should be modulated by the probabilities of these hypothetical states, known as the belief state. We hypothesized that prediction errors might be weighted by the belief state during learning under perceptual uncertainty, and that this modulation is governed by pupil-linked arousal systems. Combining pupillometry and an uncertainty-augmented reward-learning task ($N = 47$), we found that pupil responses to outcomes scaled with prediction errors and were down-weighted under higher uncertainty. This suggests that the brain's arousal systems combine newly arriving perceptual and reward information to dynamically regulate how much to learn in an uncertain world.

¹Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany; ²Humboldt-Universität zu Berlin, Berlin School of Mind and Brain, Berlin, Germany; ³Section Computational Cognitive Neuroscience, Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany ⁴Bernstein Center for Computational Neuroscience, Charité Universitätsmedizin, Berlin, Germany ⁵Max Planck Research Group NeuroCode, Max Planck Institute for Human Development, Berlin, Germany; ⁶Institute of Psychology, Universität Hamburg, Hamburg, Germany; ⁷Department of Neurology, Charité – Universitätsmedizin Berlin, Berlin, Germany; *Corresponding author: Rasmus Bruckner (rasmus.bruckner@fu-berlin.de)

Most of our decisions are guided by beliefs we acquire based on perceptual and reward information about the environment. For example, when learning whether one likes a new pastry in a bakery, one has to form an association between the type of pastry based on perceptual information and the experienced taste when eating it. Crucially, this learning process can be quite challenging when stimuli such as different sorts of pastries cannot be easily distinguished due to perceptual uncertainty, and one might learn an incorrect stimulus-reward association. Therefore, taking into account perceptual uncertainty during reward-based learning is crucial for adaptive behavior (Bruckner et al., 2020; Daw & Dayan, 2014; Ganesh et al., 2024; Ma & Jazayeri, 2014; Summerfield & Tsetsos, 2012).

One approach to addressing this issue is to combine perceptual inference with reinforcement learning. During perceptual inference, a belief state can be computed, quantifying the probability of a particular state (such as the type of pastry) based on observations. This enables a precise quantification of how well different states and stimuli in the environment can be distinguished. Subsequently, learning can operate on belief states. Once an outcome has been observed, the belief state regulates how much credit each state receives. Specifically, the belief state can be used to regulate how much one learns from a prediction error, reflecting the difference between the actual and the experienced reward (Bruckner et al., 2020; Chrisman, 1992; Ez-zizi et al., 2023; Ganesh et al., 2024; Lak et al., 2017; Larsen et al., 2010). We have recently shown that when the belief state clearly favors a state (low uncertainty), participants update their beliefs more strongly in response to a prediction error (higher learning rate). This was in contrast to situations with higher state uncertainty, in which participants learned more slowly (Ganesh et al., 2024). These results are generally in line with prior work suggesting a link between belief state and the regulation of learning from prediction errors (Babayan et al., 2018; Bruckner et al., 2020; Colizoli et al., 2018; Gershman & Uchida, 2019; Lak et al., 2017, 2020; Starkweather et al., 2017). Here, we investigate the physiological mechanisms that may support adaptive learning under perceptual uncertainty.

The arousal systems of the brainstem, such as the locus coeruleus-norepinephrine system, are in a good position to control learning under uncertainty. These systems are recruited by specific forms of uncertainty (Aston-Jones & Cohen, 2005; Dayan & Yu, 2006; A. J. Yu & Dayan, 2005), and could play a key role in the dynamic, uncertainty-dependent regulation of learning rates (Nassar et al., 2012). Through their widespread ascending projections to the cerebral cortex, they shape local and large-scale network interactions (Berridge & Waterhouse, 2003; Pfeffer et al., 2022; Podvalny et al., 2021; Van Den Brink et al., 2019), and synaptic plasticity mechanisms on different timescales (Bear & Singer, 1986; Berridge & Waterhouse, 2003; Marzo et al., 2009; Nadim & Bucher, 2014; Rasmussen, 2000; Reynolds et al., 2001; Reynolds & Wickens, 2002; Vetencourt et al., 2008) in the cortex. This potentially makes them a key neuromodulatory mechanism supporting learning under perceptual uncertainty.

The activity of brainstem arousal systems can be tracked non-invasively by measuring non-luminance-mediated changes in pupil size (De Gee et al., 2017; Joshi et al., 2016; McGinley et al., 2015; Reimer et al., 2016). Several studies show that pupil responses to outcomes encode prediction errors and surprise governing learning in uncertain and changeable environments (De Berker et al., 2016; Nassar et al., 2012; O'Reilly et al., 2013; Preuschoff, 2011; Van Slooten et al., 2018). Moreover, different forms of uncertainty relevant to learning and decision-making are reflected in pupil-linked arousal responses (Colizoli et al., 2018; De Berker et al., 2016; Krishnamurthy et al., 2017; Nassar et al., 2012; Urai et al., 2017; Van Slooten et al., 2018).

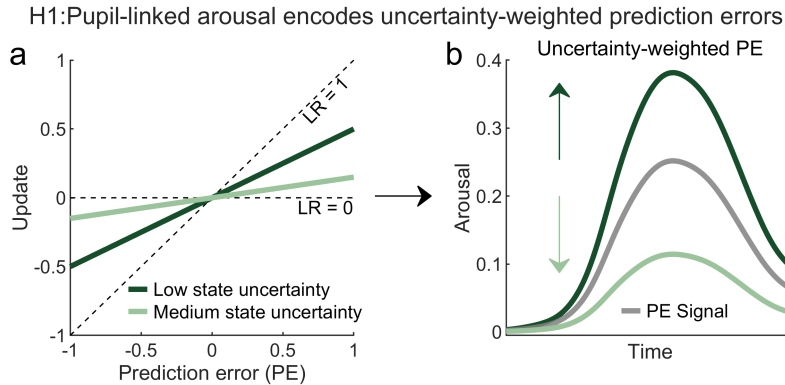


Figure 1. Arousal influences learning under perceptual uncertainty. Learning and decision-making often take place under perceptual uncertainty over stimuli or states due to ambiguous sensory information (e.g., two very similar types of pastries). In this case, a belief about the state (i.e., belief state) can be computed to quantify how confidently the states can be distinguished (e.g., 40% pastry A, 60% pastry B). Belief states that are more similar mean higher state uncertainty. In contrast, state uncertainty is lower when stimuli or states can be more clearly distinguished. **a** Consequently, under lower state uncertainty (dark green line), the learning rate (LR) tends to be higher, leading to larger updates for a given prediction error. In contrast, under higher state uncertainty (light green line), learning rates are lower. **b** We hypothesized that this adaptive learning mechanism is supported by changes in arousal due to uncertainty driving how reward information is processed and used for further belief updating. Changes in arousal levels are known to reflect a prediction-error signal (see gray curve) and are associated with uncertain environments. Here, we examine whether pupil-linked arousal encodes a prediction error that is modulated by uncertainty over the belief state. Lower levels of uncertainty are associated with an up-regulated expression of prediction error in the arousal signal (dark green curve vs. gray curve), and higher levels are linked with mitigated representation of prediction errors in the arousal level (light green curve vs. gray curve).

We built upon these findings and hypothesized that pupil responses reflect uncertainty-weighted prediction errors during reinforcement learning under perceptual uncertainty (Fig. 1). In support of our hypothesis, we show that the uncertainty reflected in a belief state modulates how strongly the pupil reacts to prediction errors. Our study suggests that arousal systems are responsive to state inference under perceptual uncertainty, regulate the computations of prediction errors accordingly, and support adaptive learning by combining perceptual and reward information through a shared neural mechanism.

Results

Task and learning behavior

To investigate if pupil-linked arousal plays a role in belief-state-guided learning, we analyzed behavioral and pupil data of 47 participants (28 female, 10 male, 9 non-binary; mean age = 25.09 ± 0.67 standard error of the mean (SEM); range: 18-40) who completed the uncertainty-augmented Gabor-Bandit task (Bruckner et al., 2020; Ganesh et al., 2024). Each trial in the task consisted of three stages that we briefly summarize here. The main aim of the task was to maximize rewards, which required that subjects accurately learn reward probabilities from feedback. In the choice phase, participants were presented with two Gabor patches and asked to choose the one that maximized rewards. In the feedback phase, they received binary reward feedback. Finally, in the slider phase, they reported their reward probability estimate with a slider ranging from 0 to 100. The reported reward probability reflected the participants' updated belief after receiving the reward feedback. This allowed us to measure learning from trial to trial (Fig. 2a). Below, we provide a more detailed explanation of the task structure.

To maximize rewards, participants had to learn the reward probability, which was governed by an underlying relationship between each trial’s state, participants’ actions, and the obtained reward (Fig. 2b). Each trial belonged to a hidden state of 0 or 1. The state determined the contrast-difference level on that trial. In state 0, the right patch had a higher contrast level compared to the left patch. In state 1, the left patch had a higher contrast level than the right one. The contrast difference between stimuli varied between trials, and participants had to choose the patch with the weaker (half of blocks) or stronger contrast (other half of blocks) to obtain the most reward. For each correct choice, participants received positive feedback with either a 90 % or 70 % probability. Positive feedback after wrong choices was given with the respective complementary probabilities (10 and 30 %, respectively). The reward probability was governed by an underlying relationship between each trial’s state, participants’ actions, and the obtained reward (Fig. 2b). Please note that participants were not aware of the mapping between the contrast of the patches and the reward probability. Thus, they were required to re-learn it in each block of trials. Finally, participants used the slider to report their reward probability estimates on a trial-by-trial basis.

We induced uncertainty over the perceptual features of the Gabor patches by manipulating contrast-difference levels on a trial-by-trial basis. The contrast differences were sampled from a uniform distribution. Belief states were more uncertain when contrast differences were smaller (patches looked more similar). In contrast, state uncertainty was lower in trials with larger contrast differences. Next to manipulating state uncertainty across trials, the task featured two levels of reward uncertainty governed by a "contingency" parameter μ . Half of the blocks had higher reward uncertainty with 70 % reward probability (e.g., contingency parameter $\mu = 0.7$) and lower reward uncertainty with 90 % reward probability ($\mu = 0.9$). Thus, the reward manipulation resulted in two task conditions (Fig. 2a inset).

To ensure that participants understood the task and learned to choose the more rewarding option, we analyzed their subjective estimate of the reward probability. We plotted mean learning curves based on participants’ slider responses, which showed that participants updated their beliefs about the reward probability in line with the feedback (Fig. 2c). Participants approached the actual underlying probability across both task conditions (high reward uncertainty: mean = 0.64 ± 0.01 , $t_{45} = 9.55$, $p < 0.001$, Cohen’s $d = 6.35$, low reward uncertainty: mean = 0.72 ± 0.02 , $t_{45} = 13.64$, $p < 0.001$, Cohen’s $d = 6.6$). There was a significant effect of task condition on the slider responses whereby the reported reward probability was significantly lower in the high reward-uncertainty condition ($t_{90} = -3.51$, $p < 0.001$, Cohen’s $d = -0.73$). Consistent with these results, participants also learned to choose the correct option (i.e., more rewarding) more often (Fig. S1). Expanding on these findings about participants’ learning behavior, we next tested whether state uncertainty impacted learning rates.

Participants used variable learning rates modulated by the belief state. Based on participants’ slider responses, we first computed each trial’s prediction error and belief update (see [Data preprocessing](#), for more details). The learning rate (LR) is then computed as the extent to which participants updated their belief about the reward probability for that trial’s prediction error. Moreover, we approximated belief states contingent on the contrast-difference level, wherein lower difference resulted in higher state uncertainty. To check if this uncertainty impacted single-trial learning rates on average, we calculated averages for increasing contrast-difference bins and found that participants learned more when belief states were less uncertain (Fig. 2d; Pearson’s $r_{08} = 0.93$, $p < 0.001$).

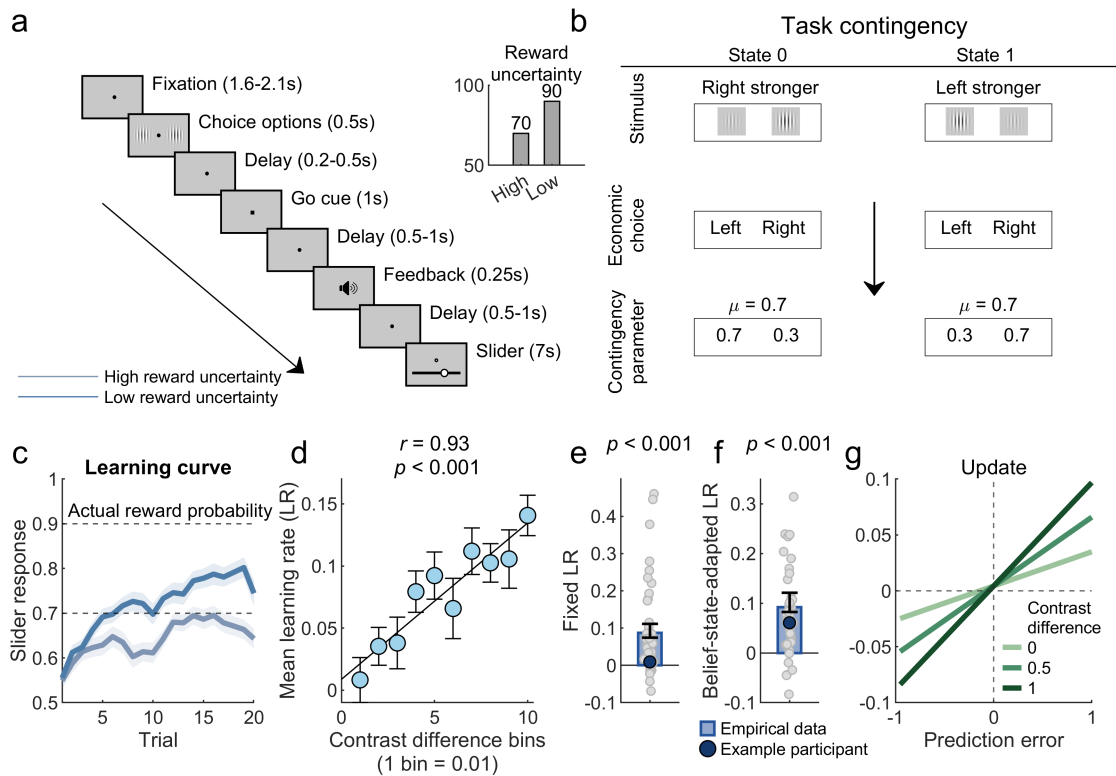


Figure 2. Experimental task and learning behavior. **a** | Uncertainty-augmented Gabor-Bandit task. Participants were instructed that the aim of the task was to maximize rewards. A pair of Gabor patches with varying contrast differences was presented on each trial. Participants made an economic choice between the patches (choosing the high- or low-contrast patch). After each choice and a jittered delay, auditory reward feedback was presented. Finally, participants reported their estimate of the reward probability (for a hypothetical choice) by using a slider between 0 (low reward probability) and 100 percent (high reward probability). **Inset plot.** We used two experimental conditions. The two conditions differed in terms of reward uncertainty. In half of the blocks, reward uncertainty was low, and when the “correct” patch was chosen, the reward probability was 90%. In contrast, reward uncertainty was higher in the other half of the blocks (70% reward probability for a correct choice). **b** | Task contingency. To maximize rewards, participants were required to learn the underlying state-action-reward contingency for each block of trials. Each trial could belong to either state 0 or 1. In state 0, the patch on the right had a higher contrast level, and vice versa in state 1. Conditional on the state, the exact contrast difference was sampled from a uniform distribution ranging from low difference values that were virtually in-discriminable (higher uncertainty) to higher values (lower uncertainty). The contingency parameter determines the probability of a reward for a given state-action combination. In this example, the reward probability is higher on choosing the left patch in state 0. Similarly, the reward probability is higher on choosing the right patch in state 1. Thus, to learn this underlying relationship, participants need to use the obtained reward to estimate the probability with which a given action is likely to be rewarded conditional on the current trial’s state. Participants face a considerable amount of uncertainty in trials with smaller LR contrast differences making state inference challenging.

Figure 2 (Continued). **c**| Learning curve. Mean \pm standard error of the mean (SEM) of the reported reward probability based on the slider plotted across trials. Across the two conditions, participants approached the actual reward probability albeit with some level of underestimation. Subjects’ reported reward probability was lower in the condition with high reward uncertainty compared to the low-reward uncertainty condition. **d**| Mean single-trial learning rates. Here, we illustrate the learning rate across contrast-difference bins, where lower bins correspond to more uncertain belief states (high state uncertainty). Subjects used higher learning rates in the larger contrast-difference bins, suggesting more learning under more distinguishable belief states. **e**| To decompose the effects of prediction errors and belief states on learning, we used a regression model (Ganesh et al., 2024). The model included a term expressing the average influence of the prediction error on the update (fixed LR). Positive coefficients show that participants have a tendency to report larger updates after larger prediction errors. **f**| The model also included a term estimating the influence of belief states on learning (belief-state-adapted LR). Positive coefficients show that participants adjust the learning rates to the level of contrast difference of a given prediction error. **g**| Regression fits for an example participant across three levels of contrast difference demonstrate that belief states modulated the learning rate. Higher contrast differences led to larger updates compared to trials with lower contrast differences.

We next decomposed single-trial belief updates to examine the extent to which the belief state is used to weigh prediction errors while learning. We used a previously developed regression model, based on which we quantified single-trial belief updates as a linear function of prediction errors (Ganesh et al., 2024). The model estimated a fixed learning rate (quantifying error-driven learning) and a belief-state-adapted learning rate (quantifying the dynamic impact of belief states in interaction with prediction errors) and controlled for the role of nuisance variables on learning (for more details, see [Behavioral regression model](#)). We fit the model to participants’ single-trial updates and found that participants used positive fixed learning rates on average (mean = 0.09 ± 0.018 , $t_{46} = 5.1$, $p < 0.001$, Cohen’s $d = 0.74$; Fig. 2e) indicating more learning from larger prediction errors. Next to this, we found that participants learned more for larger contrast differences (i.e., lower uncertainty over the belief state) for a given prediction error as indicated by the positive belief-state-adapted-LR coefficient (mean = 0.1 ± 0.019 , $t_{46} = 5.27$, $p < 0.001$, Cohen’s $d = 0.77$; Fig. 2f). We illustrate this result by plotting the relationship between prediction errors and updates for varying contrast-difference levels for an example participant (Fig. 2g), which shows a systematic increase in learning rates with lower uncertainty. Together, these results show that humans use a mixture of fixed learning driven by prediction errors and dynamic learning conditional on the belief state, replicating the results of our previous work (Ganesh et al., 2024). Finally, to quantify the performance of our model compared to a baseline model with intercept only, we computed the average *partial-R*². This yielded a *partial-R*² = 0.42, suggesting that our model captured the behavioral data well (Fig. S5). Building upon these insights, we next investigated the role of pupil-linked arousal dynamics in the adaptive learning process.

Feedback-related pupil dilation reflects prediction errors

The central question addressed in this study is whether pupil-linked arousal contributes to adaptive learning by means of encoding uncertainty-weighted prediction errors. We first quantified the arousal level as the baseline-corrected pupil signal on a given trial (Fig. 3a and Fig. S10 for single-subject pupil responses). Pupil responses during the choice phase (time-locked to patch presentation) are likely, at least in part, driven by decision formation and the associated uncertainty (De Gee et al., 2014; Urai et al., 2017). By contrast, pupil responses to feedback onset reflect learning-related computations such as surprise or the encoding of a reward prediction error (Nassar et al., 2012; Preuschoff, 2011; Van Slooten et al., 2018). We directly compared pupil responses locked to feedback onset (for details see, [Quantifying feedback-related pupil signal](#)) across two bins of absolute prediction errors and found that pupil dilation was significantly

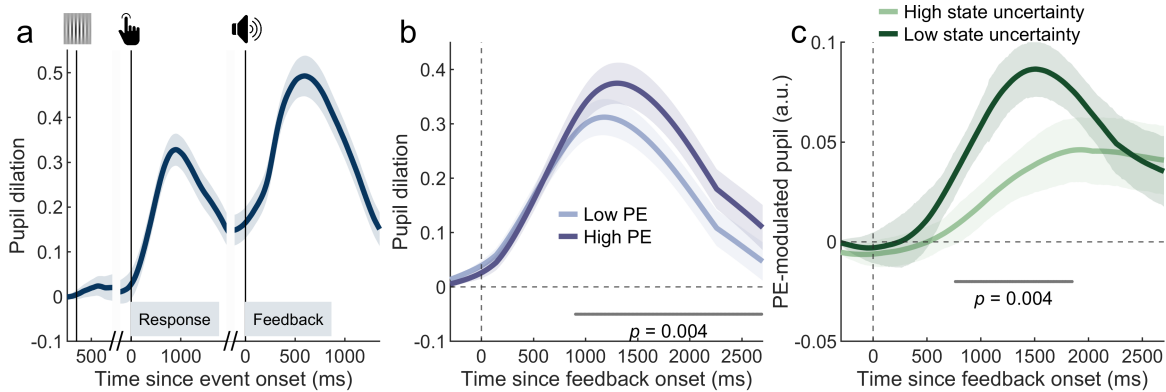


Figure 3. Feedback-locked pupil dilation. **a**| Baseline-corrected time course of mean \pm standard error of the mean (SEM) pupil dilation throughout the trial. The positive response post feedback onset suggests the encoding of a reward prediction error, while patch and response-locked dilation reflects task-based decision-making processes. **b**| Mean pupil response to feedback onset was significantly larger for trials with higher compared to lower prediction errors. **c**| Prediction error-related pupil dilation for trials with low and high state uncertainty. To examine if state uncertainty modulates the extent to which prediction errors scaled pupil dilation, we regressed pupil dilation onto absolute prediction errors for each time point post feedback onset for high and low state-uncertainty trials separately. Regression coefficients were significantly larger for lower state-uncertainty trials compared to trials with higher state uncertainty. This result suggests that prediction errors are down-regulated as a function of state uncertainty. The gray line indicates p -values from a cluster-based permutation test of the difference between two groups of trials, after accounting for multiple comparisons.

larger for larger prediction errors (dark purple curve in Fig. 3b, $p = 0.004$, see [Permutation testing](#) for more details). This finding is in line with the idea that larger errors signal stronger surprises.

We then examined if the prediction-error component of the pupil response is modulated by the belief state. To this end, we used a model-based approach whereby we fit absolute prediction errors and other control variables (see [Pupil regression analysis](#) for more details on the control variables) to individual subjects' pupil dilation in a sample-by-sample fashion across two bins of high and low state uncertainty (Urai et al., 2017; see [Binned regression approach](#) and Fig. S2 for more details on the binned regression approach). By comparing beta coefficients that quantified the extent to which prediction errors modulated pupil responses, we found that the relationship between absolute prediction errors and pupil dilation is systematically down-regulated for higher uncertainty over the belief state (light green curve in Fig. 3c, $p = 0.004$, see [Permutation testing](#) for more details and Fig. S3 showing that the mean absolute prediction errors were not significantly different across the two bins). This potentially reveals uncertainty-weighted prediction errors, where higher state uncertainty calls for lower levels of error-evoked arousal. To examine this idea in more detail, we next turn to an analysis approach with continuous levels of belief states.

Pupil signal reveals uncertainty-weighted prediction errors

In a model-based regression analysis, we found further evidence of uncertainty-weighted prediction errors (Fig. 4). While our previous analysis approximated this relationship by comparing high- and low-uncertainty trials (Fig. 3), our next analysis directly tested for the relationship between prediction errors and contrast differences. We used a linear regression model for each subject across the time course of feedback-locked pupil dilation (Krishnamurthy et al., 2017; Nassar, Bruckner, & Frank, 2019; Urai et al., 2017; Van Slooten et al., 2018). The model estimates the temporal evolution of the extent to which pupil dilation changes for different re-

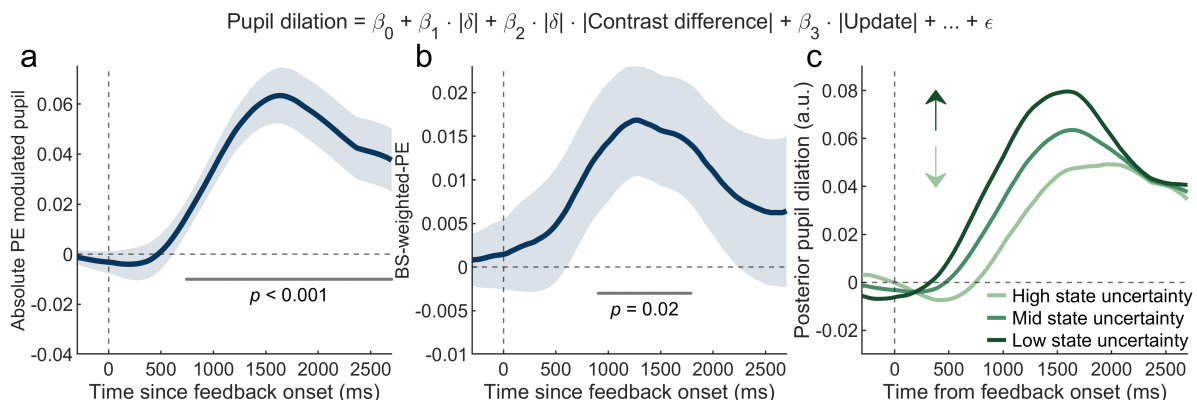


Figure 4. Feedback-locked pupil signal encodes belief-state-weighted prediction errors. **a**) We developed a regression model (see inset equation on top of the plot) and fit it to sample-by-sample pupil response to feedback. Positive coefficients for absolute prediction errors $|\delta|$ show a systematic increase in the pupil response for larger prediction errors showing that prediction errors are reflected in the pupil response. **b**) Positive coefficients for the interaction between absolute prediction errors and contrast differences (i.e., belief-state-weighted prediction errors) indicate that pupil responses increased significantly for larger contrast-difference trials (i.e., lower state uncertainty) compared to smaller contrast differences for a given level of prediction error. This finding suggests that prediction errors are modulated by the belief state. **c**) Across three levels of contrast-difference values, model-predicted responses for a given prediction error demonstrate that state uncertainty plays a key role in regulating the pupil response. Higher contrast differences (i.e., on average, lower state uncertainty) led to increased pupil dilation compared to trials with lower contrast differences. The gray line indicates p -values from a cluster-based permutation test of the difference between each time course and zero.

gressors. We first accounted for the impact of absolute prediction errors by adding a main effect to the model. To model the modulatory role of the belief state, we added a key regressor that captured the interaction effect between contrast difference and absolute prediction error on a given trial (Fig. 4 inset equation, where $|\delta|$ refers to absolute prediction error). We also added a set of nuisance regressors to simultaneously control for the impact of the main effect of the belief state, reaction times, slider responses, and gaze position (for more details on the model, see [Pupil regression analysis](#) and Fig. S4 for the effect of control variables on the pupil response).

We found that pupil dilation varied systematically as a function of absolute prediction errors. That is, larger prediction errors resulted in a significant increase in the magnitude of pupil dilation post feedback (PE-modulated pupil coefficient in Fig. 4a, $p < 0.001$, see [Permutation testing](#) for more details). The positive influence of prediction errors reflects an arousal-based signal for error-driven learning dynamics reminiscent of the behavioral fixed learning rate (Fig. 2e).

More importantly, we determined that uncertainty over the belief state modulates the pupil responses to prediction errors during learning. Positive coefficients for the interaction of absolute prediction error and contrast difference indicated that larger contrast differences (more certain belief states) led to increased pupil dilation for a given prediction error (BS-weighted PE coefficient in Fig. 4b, $p = 0.02$, see [Permutation testing](#) for more details). In other words, for a given prediction error, pupil dilation is attenuated on trials with higher state uncertainty and enhanced with decreasing uncertainty levels. This suggests that via regulated arousal, state uncertainty flexibly changes how incoming reward information is translated into learning. This result qualitatively complements the behavioral adaptive learning rate (Fig. 2f).

To demonstrate prediction-error weighting in the arousal signal further, we computed feedback-locked pupil dilation based on our model-estimated coefficients (Fig. 4c). Mean posterior dilation plotted across three contrast-difference values showed that lower model-predicted pupil dilation accompanied higher levels of uncertainty (smaller contrast-difference values). Taken

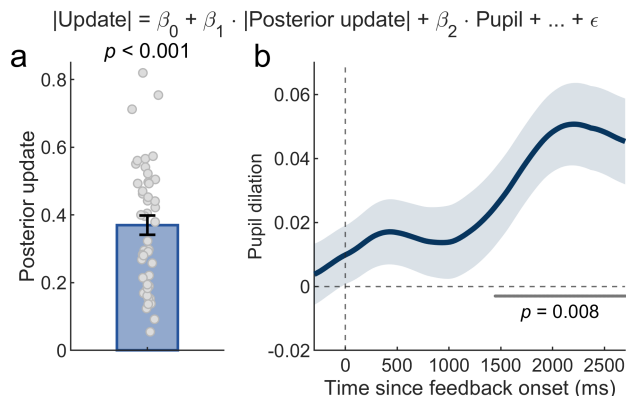


Figure 5. Pupil-linked arousal explains belief updating beyond our behavioral model. **a**| We examined the relationship between pupil-linked arousal and residual learning, which captures additional variance in behavior after having accounted for the effects explained by our behavioral regression model (see the inset equation on top of the plot). We include the posterior-update coefficient that summarizes the model-based behavioral predictions, allowing us to control for variance in updates explained by the behavioral regression model. **b**| Feedback-locked pupil dilation significantly contributed to explaining residual learning from our behavioral model. This shows that fluctuations in pupil-linked arousal levels explain variance in learning behavior beyond the behavioral regression model. This suggests that higher arousal levels are related to larger belief updates, potentially because newly arriving information is considered more strongly than under lower levels of arousal. The gray line indicates p -values from a cluster-based permutation test of the difference between each time course and zero.

together, our finding suggests that pupil dilation reflects an uncertainty-weighted prediction error that could drive adaptive learning. Please note that we also used versions of our current regression model to analyze the pupil signal after controlling for the role of baseline pupil signal and reaction times. Results indicated that the estimated coefficients converge with our current results (Fig. S7 and Fig. S8; see [Control regression models](#) for more details).

More arousal is linked to larger belief updates

Finally, we directly examined the link between pupil responses and behavioral updating. Based on our findings, we reasoned that since dynamic information processing is key for adaptive learning and is tightly associated with arousal levels, fluctuations in pupil dilation may be related to differences in update magnitude. To test this, we investigated whether variance in absolute single-trial updates has a systematic relationship with pupil-linked arousal after accounting for key regressors from our behavioral model (“posterior updates”). We applied a regression model explaining belief updating, but in this case, we directly took into account pupil dilation in the model. Moreover, by adding posterior updates as regressors (Fig. 5a, $p < 0.001$), our analysis ensured that pupil-linked arousal is used to solely explain residual learning behavior beyond our behavioral model. Applying this model, we found that arousal levels modulate belief updating beyond the variance captured in our behavioral analyses. In this model, the main effect of pupil dilation represented the direct relationship between arousal and updates (for more details, see [Analyzing residual learning](#)). Larger pupil dilation in the later periods post feedback co-varied positively with participants’ single-trial updates (Fig. 5b, $p = 0.008$, see [Permutation testing](#) for more details). This suggests that higher general levels of arousal increase update magnitude so that newly incoming information is more strongly considered for learning. Residual learning is better explained by such fluctuations in the arousal level that occur later during the learning window (Fig. 5b). In sum, pupil dilation not only reflects task-related learning computations but also intrinsic arousal fluctuations, which explain learning beyond those captured by our behavioral model.

Discussion

Recent work has indicated that the arousal systems of the brainstem may play an important role in controlling adaptive behavior in an uncertain world. Here, we examined whether phasic pupil-linked arousal supports adaptive learning under perceptual uncertainty. Our results show that pupil responses during learning are flexibly modulated by uncertainty over stimulus discriminability (belief state) while encoding prediction errors. This suggests that pupil-linked arousal represents prediction errors weighted by the belief state. The dynamic regulation of prediction errors might help the brain to flexibly adjust reward learning to varying degrees of uncertainty due to ambiguous perceptual information.

Our findings suggest that phasic pupil-linked arousal encodes an uncertainty-weighted prediction-error signal. Accordingly, the pupil response reflects computations that integrate perceptual information encoded by the belief state with reward information through prediction errors. Such an uncertainty-mediated arousal signal flexibly changes how incoming reward information is treated during learning. This could reveal an adaptive computational mechanism for slower and more cautious learning when stimuli cannot be confidently dissociated in the presence of perceptual uncertainty (Bruckner et al., 2020; Drevet et al., 2022; Ganesh et al., 2024).

While our results are the first to demonstrate that phasic pupil-linked arousal encodes uncertainty-weighted prediction errors during adaptive learning, previous work has hinted that this might be the case. Several studies reported pupil correlates of prediction errors and surprise, independent of perceptual uncertainty (Browning et al., 2015; De Berker et al., 2016; Nassar et al., 2012; O'Reilly et al., 2013; Preuschoff, 2011; Van Slooten et al., 2018). The relationship between belief states and prediction errors is consistent with a perceptual decision-making study that interpreted outcome-locked pupil responses in terms of an uncertainty-weighted prediction error (Colizoli et al., 2018). Finally, previous animal work measuring dopaminergic activity showed evidence of uncertainty-weighted prediction errors in mice and monkeys (Lak et al., 2017, 2020).

More generally, our results add to the current view that state inference is a ubiquitous mechanism for adaptive learning under uncertainty. Recent studies have also shown prediction-error weighting in tasks in which uncertainty arises from ambiguous reward information (Babayan et al., 2018; Starkweather et al., 2017). Together with our findings, this suggests that reinforcement learning operates on belief states when the states of the environment are only partially observable. Specifically, learning under uncertainty involves two key processes: first, the brain computes the current belief state based on the available perceptual information, which then serves as input for the reinforcement-learning system (Daw & Dayan, 2014; Dayan & Daw, 2008; Wilson et al., 2014). In the case of our study, this is reflected in an uncertainty-weighted prediction error signal. This dynamic between state inference and learning seems to function consistently, regardless of whether the source of uncertainty primarily stems from sensory information, reward information, or other sources.

Further pupillometry research could expand on these considerations and examine whether task-evoked arousal encodes uncertainty-weighted prediction errors when uncertainty primarily arises from ambiguous reward magnitudes or timing. Moreover, future studies could examine where in the brain belief states are computed and passed on to regulate learning and decision-making. Plausible candidate areas are the orbitofrontal cortex (OFC), hippocampus, and entorhinal cortex, wherein neuromodulatory surprise signals can mediate state inference (L. Q. Yu et al., 2021). Theoretical and empirical work suggests that the OFC represents hidden-state information (Nassar, McGuire, et al., 2019; Schuck et al., 2016; Wilson et al., 2014), and it will be important to better understand whether the OFC also represents state uncertainty that modulates prediction errors for reinforcement learning.

Our findings potentially involve key neuromodulatory systems. Mounting evidence identifies the locus coeruleus-norepinephrine (LC-NE) system, a core component of the arousal network, as a likely contributor to phasic pupil responses, which influences the central arousal state through its extensive and rich cortical projections (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010; Joshi & Gold, 2020; Joshi et al., 2016; Megemont et al., 2022; Murphy et al., 2011, 2014; Reimer et al., 2016). Acetylcholine has also been related to LC activity, which is quantified as changes in both tonic (Reimer et al., 2016) and task-evoked pupil responses (Sara, 1998). This is in line with influential theoretical work proposing that norepinephrine and acetylcholine track different forms of uncertainty that regulate learning behavior (Dayan & Yu, 2006; Sara, 1998; A. J. Yu & Dayan, 2005). Moreover, the dopaminergic system, known for its role in reward- and error-related computations and with connections to the LC, has been linked to changes in pupil diameter (De Gee et al., 2017; Manohar & Husain, 2015). Other serotonergic regions and the superior and inferior colliculi have also been tied to pupil-linked arousal (Cazettes et al., 2021; Joshi et al., 2016; Reimer et al., 2016). This suggests a complex interplay between different neuromodulatory systems in regulating arousal. In humans, future work using functional magnetic resonance imaging (fMRI) data from a wider brainstem network (De Gee et al., 2017) could help identify the neural source behind uncertainty-regulated pupil responses and its link to learning-rate adjustments.

We next discuss the potential role of cognitive effort and the relationship to uncertainty-weighted prediction errors. Because pupil dilation could reflect cognitive effort, one potential interpretation of our results is that lower prediction-error signals on high-uncertainty trials could be explained by the expenditure of higher effort on such trials. However, at least two arguments speak against this interpretation. First, we have included contrast differences as main effects in our pupil analyses. If participants invested higher cognitive effort on high-uncertainty trials (smaller contrast differences), we would expect significant main effects of contrast differences. However, neither during the presentation of the Gabor patches (Fig. S9) nor in response to reward feedback (Fig. S4) did we find significant main effects of contrast difference. Second, previous studies have found that pupil responses are larger in high-effort trials (Van Der Wel & Van Steenbergen, 2018). In contrast, the uncertainty-weighted prediction-error signal shows that more uncertainty (hence potentially more effort) calls for lower pupil responses. Therefore, our findings seem inconsistent with the interpretation that down-weighted prediction-error signals for smaller contrast differences reflect the investment of more effort.

Finally, our analysis also showed that in addition to encoding uncertainty, fluctuations in the phasic pupil response capture variability in learning beyond what could be explained through behavioral analyses alone. This indicates that pupil-linked arousal not only reflects task-based computations but also accounts for fluctuations in learning. This finding is in line with the study by De Gee et al. (2017) showing that seemingly random variability in choice behavior can partly be explained by phasic responses of brainstem arousal. Together, these lines of research converge on the view that arousal is a ubiquitous factor that exercises control over how the brain processes information for learning and decision-making.

In conclusion, we found support for the hypothesis that phasic pupil-linked arousal reflects prediction errors modulated by belief states during reinforcement learning under perceptual uncertainty. This suggests that the arousal systems are key in the integration of uncertain perceptual and reward information and offers new insights into the psychophysiological mechanism behind adaptive belief updating.

Methods

Participants

We collected behavioral and pupillometry data from 50 participants (29 female, 11 male, 10 non-binary; mean age = 25.59 ± 0.6 SEM; range: 18-40). All participants were recruited from the local student participant pool of Freie Universität Berlin. Data from 3 participants were rejected since they performed with less than 50% accuracy. Thus, all further analyses were conducted on the data obtained from our effective study sample of 47 participants. For taking part in the study, participants were paid a standard rate of 12.00 €. Moreover, to incentivize their performance, participants received an extra bonus payment of up to 4.00 €, determined by their economic choice performance. The study was approved by the ethics committee of the Department of Education and Psychology at Freie Universität Berlin ("Effects of Perceptual Uncertainty on Value-Based Decision Making", protocol number: 121/2016). All participants reported normal or corrected-to-normal vision, seven reported currently being diagnosed with a psychiatric condition (e.g., depression, anxiety, schizophrenia) and none reported being diagnosed with a neurological condition (e.g., multiple sclerosis).

Experimental task

We used an isoluminant version of the Gabor-Bandit task from Ganesh et al. (2024) optimized for pupillometry.

Stimulus material

The task was programmed using `PsychoPy` (version 2022.2.5). The Gabor-Bandit task version of this study comprised three stages (economic decision-making, reward feedback, slider response) (Fig. 2a). In the first stage, the stimulus material comprised a fixation dot and two Gabor patches presented on a screen with a gray background color (#808080). We set the stimulus properties of the Gabor patches using the in-built `GratingStim` function of `PsychoPy`. We used a sine texture with a spatial frequency of 8 cycles in height and a Gaussian mask. We set the orientation and phase offset of the Gabor-patch stimuli to 0° . To manipulate the contrast levels of the Gabor patches g , we adjusted the patches' visibility v , where 0 indicates that the patch is transparent (matching the background) and 1 that it is fully opaque. Consequently, the presented contrast level of each patch was a weighted combination of the stimulus properties z (as described above) and the background color h : $g = vz + (1 - v)h$. The mean visibility of both patches was kept at $v = 0.5$ across all trials (the average contrast value on each trial was 0.5 irrespective of the contrast difference). The mean contrast was constant across all trials, and Gabor-patch pairs were matched on luminance across all trials. This was verified using the `SHINE` toolbox in *Matlab* (Willenbockel et al., 2010) whereby pairs of Gabor patches with varying levels of contrast differences were tested to check if they matched on luminance levels. The stimuli was presented on a monitor with a refresh rate of 60 Hz and a length of 60.5 cm.

Experimental procedure

At the start of each trial, participants were asked to fixate on the fixation dot for a jittered time interval between 1.6 to 2.1 seconds. Participants were instructed to fixate on the dot throughout the trial and avoid blinking or making eye movements as much as possible. In the first stage of the trial, the choice gratings were presented for 0.5 seconds, while the fixation dot remained on the screen throughout. Following a jittered delay of 0.2 to 0.5 seconds, the fixation dot would turn into a square (Go cue), and participants were required to make the economic choice using

the left and right cursor buttons of the computer keyboard. The diameter of the fixation dot and the length of the squared Go cue were matched across all trials. This ensured that there were no sudden shifts in the area of the screen covered by the fixation dot or the Go cue.

In the second trial stage after a variable delay of 0.5 to 1 second, participants were presented with audio feedback of winning either zero (descending tone from 950 to 350 Hz) or one (ascending tone from 350 to 950 Hz) point based on their economic choice for 0.25 seconds. In case of a missed trial, we used white noise as audio feedback. Please note that we rejected missed trials from our behavioral and pupillometry analysis. Participants were pre-trained to learn the association between the audio feedback and the number of points. Finally, in the third trial stage, the task included an additional probe phase. In this phase, participants were first presented with a slider to report their learned estimate of the reward probability after a delay of 1.5 seconds. Participants were instructed to use the delay to estimate how much they wanted to update their beliefs after the reward feedback. This allowed us to also measure the pupil signal in the "learning" window independent of movements made by the participant to make a slider response. Finally, participants reported their subjective estimate of the reward probability for a hypothetical choice using a slider. Participants completed 20 trials in each block of the task. The presentation order of blocks was randomized across participants. If a participant failed to respond to a trial, the same trial was repeated at the end of the block.

Task contingencies

Participants were instructed to learn the underlying state-action-reward association to maximize their rewards. One of the two hidden tasks states $s_t \in \{0, 1\}$ would determine the contrast difference on each trial. For instance, if the state of a given trial is 0 ($s_t = 0$), the contrast difference would be negative. That is, the patch on the left side of the fixation dot would have a lower contrast level than the right patch. This relationship is reversed when the trial belongs to $s_t = 1$. Each block is evenly split, with half of the trials in $s_t = 0$ and the other half in $s_t = 1$. The trial's state, in conjunction with the participant's choice (left vs. right patch, referred to as actions $a_t \in \{0, 1\}$), would determine the reward probability. For example, if the participant chose the left patch ($a_t = 0$), which is the lower contrast option in $s_t = 0$, the reward probability would be higher. Similarly, when in state $s_t = 1$ and choosing action $a_t = 1$, which is the lower contrast option, the reward probability would be higher. These blocks are known as the low-contrast blocks. Crucially, in half of the blocks, the state-action-reward contingency was reversed i.e., the other half of the blocks were known as high-contrast blocks where the higher contrast options were more rewarding ($s_t = 0$ and $a_t = 1$ or $s_t = 1$ and $a_t = 0$). Participants were required to relearn the reward contingencies on each block since the contingency could be different from one block to the next due to the randomized presentation of blocks.

Task details

The main task comprised 8 blocks with 20 trials across two conditions: the high-reward uncertainty condition and the low-reward uncertainty condition. In the high-reward uncertainty condition, the reward probability was 70 %, and in the low-reward uncertainty condition, the reward probability was 90 %. Across both conditions and for each trial, the contrast difference between the two patches was randomly sampled from a uniform distribution of $[-0.1$ to $0]$ when the trial belonged to $s_t = 0$ and $[0$ to $0.1]$ for trials where $s_t = 1$. Thus, the absolute contrast levels of the patches ranged from 0.40 to 0.60.

During the slider probe phase, the contrast difference was not manipulated across trials, and the patches were clearly distinguishable. To report the estimated reward probabilities on each trial, participants were provided with a slider that ranged from 0 to 100%. To prevent

participants from relying solely on the more rewarding option as the hypothetical choice to estimate the state-action-reward relationship, we manipulated the hypothetical choice across task blocks. In half of the blocks, the hypothetical choice during the slider phase was congruent with the more rewarding option in the given block (congruent blocks). However, on the other half of the blocks, the hypothetical choice was incongruent with the more rewarding patch in that block of trials (incongruent blocks). Finally, the order of task blocks was randomized for each participant to avoid any order effects.

Gabor-Bandit task model

To describe our behavioral analysis, we first present a model of the Gabor-Bandit task introducing the key parameters. In line with Bruckner et al. (2020),

- $T := 20$ stands for the number of trials per block, where we use t to indicate the trial index in all subsequent analyses,
- $S \in \{0, 1\}$ denotes the set of task states, where 0 indicates that the patch on the right has a stronger contrast level than the left patch and vice versa for state 1; the state also determines the action-reward contingency in the task,
- $C \in [-\kappa, \kappa]$ is the set of contrast differences between the patches, where κ indicates the highest contrast difference,
- $A \in \{0, 1\}$ refers to the set of economic choices, where 0 refers to choosing the left patch, and 1 refers to choosing the right patch,
- $R \in \{0, 1\}$ indicates the set of rewards,
- $p^\phi(s_t)$ is the Bernoulli state distribution defined by

$$p^\phi(s_t) := B(s_t; \phi) \quad (1)$$

with $\phi := 0.5$, which is the probability for each state,

- $p(c_t|s_t)$ is the state-conditional contrast-difference distribution defined by the uniform distribution

$$p^\kappa(c_t|s_t) := U(c_t; [-\kappa, 0])^{1-s_t} U(c_t;]0, \kappa])^{s_t} \quad (2)$$

- $p^{a_t, \mu}(r_t|s_t)$ is the action-, contingency-parameter, and state-conditional reward distribution. This distribution is defined by

$$p^{a_t, \mu}(r_t|s_t) := (B(r_t; \mu)^{1-s_t} B(r_t; 1 - \mu)^{s_t})^{1-a_t} (B(r_t; 1 - \mu)^{1-s_t} B(r_t; \mu)^{s_t})^{a_t} \quad (3)$$

with contingency parameter $\mu := 0.9$ for half of the blocks and $\mu := 0.1$ for the other half under lower reward uncertainty. Similarly, the contingency parameter $\mu := 0.7$ for half of the blocks and $\mu := 0.3$ for the other half under higher reward uncertainty.

Data preprocessing

For our statistical analyses, we relied on participants' single-trial slider responses, from which we derived updates, prediction errors, and learning rates (Ganesh et al., 2024).

- $\hat{\mu}_t$ denotes the subject's slider response, which is the subject's estimate of the contingency parameter μ_t in the Gabor-Bandit task. Please note that half of the blocks were congruent (the subject was asked to use the slider to report their estimate of the contingency parameter of the "correct", i.e., the more rewarding option). The other half of the blocks were incongruent (the subject was asked to use the slider to report their estimate of the contingency parameter of the "incorrect", i.e., the less rewarding option). To map the slider responses from both the congruent and incongruent blocks onto a common scale, we re-coded the responses of incongruent blocks according to

$$\hat{\mu}_t = 1 - \hat{\mu}_t \quad (4)$$

- $Q \in \{0, 1\}$ indicates a correct ($q = 1$) and incorrect choice ($q = 0$), defined by

$$q = \begin{cases} 1, & s = 0 \wedge a = 0 \wedge \mu > 0.5 \\ 1, & s = 0 \wedge a = 1 \wedge \mu < 0.5 \\ 1, & s = 1 \wedge a = 1 \wedge \mu > 0.5 \\ 1, & s = 1 \wedge a = 0 \wedge \mu < 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

- $D \in [-1, 1]$ stands for the set of prediction errors, defined by

$$\delta_t = \begin{cases} (\tilde{r}_t - \hat{\mu}_t), & \pi_0 \geq \pi_1 \\ ((1 - \tilde{r}_t) - \hat{\mu}_t), & \pi_0 < \pi_1 \end{cases} \quad (6)$$

where $\tilde{r}_t := r_t + a_t(-1)^{2+r_t}$. That is, to compute the prediction error, we take into account the state-action-reward contingency defined in the task model (eq. (3)). For example, when a trial's presented contrast difference favors $s_t = 0$, we assume $\pi_0 > \pi_1$. Conditional on action $a_t = 0$, the expected reward probability is $\hat{\mu}$. To account for the action dependency of the reward, we compute \tilde{r}_t from r_t , so that, for example, $r_t = 0$ conditional on $a_t = 1$ corresponds to $\tilde{r}_t = 1$ re-coded for action $a_t = 0$ (where $r_t = 1$ had it been chosen). Similarly, to account for the state dependency of the reward, we rely on $(1 - \tilde{r}_t)$ when state $s_t = 1$ is more likely than $s_t = 0$,

- $U \in [-1, 1]$ denotes the set of updates, defined by

$$u_t = \hat{\mu}_t - \hat{\mu}_{t-1} \quad (7)$$

- $B \in \{0, 1\}$ indicates a choice-confirming outcome ($b_t = 1$) and a choice-disconfirming outcome ($b_t = 0$), defined by

$$b_t = \begin{cases} r_t, & s = 0 \wedge a = 0 \wedge \mu > 0.5 \\ r_t, & s = 0 \wedge a = 1 \wedge \mu < 0.5 \\ r_t, & s = 1 \wedge a = 1 \wedge \mu > 0.5 \\ r_t, & s = 1 \wedge a = 0 \wedge \mu < 0.5 \\ 1 - r_t, & \text{otherwise} \end{cases} \quad (8)$$

- $K \in \{0, 1\}$ denotes the set of congruence-trial types, where $k_t = 0$ denotes an incongruent and $k_t = 1$ a congruent trial type,
- $L \in \{0, 1\}$ denotes the set of contrast-trial types, where $l_t = 0$ denotes a low-contrast and $l_t = 1$ a high-contrast trial,
- m_t represents log-transformed reaction times obtained from the economic choice phase.

Modeling learning behavior

Signed learning-rate analysis To tease apart the factors influencing the single-trial updates, we developed a regression model that allowed us to dissociate multiple factors driving the learning rate (Ganesh et al., 2024):

$$u_t = \beta_0 + \beta_1 \cdot \delta_t + \beta_2 \cdot \delta_t \cdot |c_t| + \beta_3 \cdot \delta_t \cdot [b_t = 1] + \beta_4 \cdot \delta_t \cdot [k_t = 1] + \beta_5 \cdot \delta_t \cdot [l_t = 1] \quad (9)$$

In this model, β_0 represents the intercept. The β_1 coefficient models the average effect of the prediction error on the update, which we interpret as the fixed learning rate (LR), commonly denoted as α in reinforcement learning. To capture how participants adjusted their learning based on belief states arising from different levels of contrast differences, we included an interaction term between prediction error and absolute contrast difference. We refer to this coefficient β_2 as the belief-state-adapted LR. Next, to control for the effect of the confirmation bias on learning, we use the interaction term β_3 between prediction error and whether an outcome is choice confirming. This is coded as a categorical variable, i.e., 0 for outcomes that disconfirm the choice and 1 for outcomes that confirm the choice. Additionally, we added two task-based block-level categorical variables as control regressors. β_4 was the interaction term between contrast (high vs. low contrast blocks) and prediction error where 0 denoted trials in a low-contrast block and 1 for trials in a high-contrast block, and β_5 captured effects of congruence (congruent vs. incongruent block type) in interaction with prediction error where 0 denoted trials in an incongruent block and 1 for trials in a congruent block. Please note that we did not put any constraints on the range of predicted updates.

Absolute learning-rate analysis Our signed learning-rate approach estimates coefficients that could represent (i) uncertainty-driven calibration of learning or (ii) learning corrupted due to state confusion under uncertainty. We performed an absolute learning-rate analysis approach to tease these two apart and quantify how uncertainty impacts update magnitude. We used absolute prediction errors $|\delta_t|$ and updates $|u_t|$ in conjunction with all the other task-based regressors mentioned in eq. (9). The coefficients estimated from this analysis reflect the magnitude of belief updates for various factors independent of whether participants updated the slider in the correct or incorrect direction. Next, we used these estimated coefficients to compute our current regression model’s predicted absolute updates. We combined single-trial task-based regressor values with the estimated parameters to compute absolute posterior updates p_t . Thus, the posterior updates summarize the model-based predictions about learning behavior.

Model properties Across both sets of analyses, all continuous regressors were re-scaled within the range of 0 and 1 by means of the min-max normalization method

$$x_t^* = \frac{(x_t) - \min(X)}{\max(X) - \min(X)} \quad (10)$$

where X is the variable of interest, x_t is the value on a given trial that gets normalized, and x_t^* is the normalized value for a given trial. This was not done for prediction errors, and we used its natural scale since it was key to retain its original sign for the signed LR analyses. We excluded trials where $\delta_t = 0$ since these trials do not necessitate belief updates. Moreover, using a canonical linear regression model is based on the assumption that the residuals are homoscedastic, that is, similar across the range of the predictor variable. However, in our model, the assumption of homoscedasticity is violated, particularly for larger prediction errors. Thus, we accounted for heteroscedasticity by using a weighted regression model, wherein more weight is given to the observations with smaller residuals providing more reliable information.

Regression diagnostics We computed a partial- R^2 to capture the extent to which our signed and absolute behavioral regression models explain learning behavior relative to a baseline model with no task-based predictors. To do so, we used the following approach

$$\text{partial-}R^2 = \frac{SSE_{baseline} - SSE_{behv}}{SSE_{baseline}} \quad (11)$$

where SSE_{behv} is the sum of squared errors of our signed or absolute behavioral regression model and $SSE_{baseline}$ is the sum of squared errors of a baseline regression model with an intercept and excluding all other predictors.

Pupillometry

Pupil diameter along with horizontal and vertical eye position was measured from the left eye using an EyeLink 1000 plus desktop mount eye-tracker (SR Research, Ontario, Canada) at a sampling rate of 1000 Hz while participants performed the behavioral task in a dark room. Participants were instructed to place their heads on a chin rest during the task. The distance between the screen and the chin rest was kept constant at 60 cm. Calibration and validation procedures were performed before the start of the task. A calibration was accepted only if the eye tracker recognized it as "stable". After the calibration procedure was completed, we performed validation. If the deviation between the actual and predicted gaze location based on calibration was more than 4 mm of visual angle, then the eye-tracker was re-calibrated. The procedures were repeated between blocks if the participants moved their heads away from the chinrest. A nine-point calibration grid was used wherein each calibration target was a circle that was black in color (#000000).

Preprocessing We preprocessed the raw pupil signal for all further analyses using a custom *Matlab* preprocessing pipeline. First, blinks that resulted in missing pupil data were identified automatically by the in-built EyeLink software. A time window of 150 ms was used to add padding before and after the blink on which linear interpolation was applied. We also used peak detection on the velocity of the pupil signal to find additional blinks which were also linearly interpolated (Urai et al., 2017). Pupil data were then band-pass filtered using a third-order Butterworth filter with a pass band of (0.01, 10 Hz) (Knapen et al., 2016). Low-pass filtering removes measurement noise from the signal. High-pass filtering removes meaningful slow drift from the signal. To do so, we applied a low-pass filter, and the filter coefficients acted as the inputs to the band-pass filter. Moreover, we applied a high-pass filter, and the band-pass filter itself was run over to the high-pass-filtered data. These steps were based on the in-built *Matlab* `butter` and `filtfilt` functions. Next, to get rid of dips in the pupil signal that typically follow a blink or a saccade, we used the filtered pupil signal to first estimate the response to blinks and saccades using deconvolution. Once these responses were identified, they were regressed out of the pupil signal (Knapen et al., 2016). Finally, the slow drift obtained from the low-pass-filtered signal was added back to the cleaned pupil signal. The pupil signal was then z-scored for each participant and down-sampled to 100 Hz.

Pupil data analysis

Quantifying feedback-related pupil dilation To examine our questions specific to the influence of state uncertainty on arousal levels during learning, we first obtained pupil dilation for each trial after the feedback presentation. For this purpose, we relied on the pupil diameter

on each trial after the onset of feedback until the end of the trial. Next, we baseline-corrected it by subtracting the mean pupil diameter for 300 ms before the event onset to obtain each trial’s feedback-related phasic pupil dilation. For all model-based analyses, we used the computed single-trial pupil dilation.

Modeling feedback-related pupil dilation To examine the relationship between task-based and single-trial behavioral parameters and baseline corrected pupil response to feedback onset, we applied the following regression model to single-trial feedback-related pupil dilation for each time point:

$$\begin{aligned} \text{Pupil} = & \beta_0 + \beta_1 \cdot |\delta_t| + \beta_2 \cdot |\delta_t| \cdot |c_t| + \beta_3 \cdot |c_t| \\ & + \beta_4 \cdot m_t + \beta_5 \cdot |u_t| + \beta_6 \cdot x_t + \beta_7 \cdot y_t \end{aligned} \quad (12)$$

β_0 is the intercept. β_1 is the coefficient that reflects the main effect of absolute single-trial prediction errors on pupil dilation. We refer to this as the prediction-error-modulated pupil coefficient as it models the impact of prediction error magnitude on pupil dilation. β_2 is the coefficient that captures the interaction effect between absolute prediction errors and contrast-difference levels. This regressor approximates the changes in pupil dilation for different levels of state uncertainty for a given prediction error. We refer to this as the belief-state-weighted prediction error coefficient. Next, to check if state uncertainty had a direct influence on pupil dilation, we added the main effect of contrast difference as a regressor (β_3). Next, we added the following nuisance variables as control regressors. β_4 captures variance explained by reaction time. We also added absolute single-trial belief updates as a regressor (β_5) to control for variance in pupil dilation due to larger slider responses and its collinear relationship with absolute prediction errors. Next, to control for any fluctuations in pupil dilation resulting from changes in horizontal (x_t) and vertical (y_t) eye position, we added regressors β_6 and β_7 , respectively. All variables were z-scored separately for each participant.

Binned regression approach To examine if state uncertainty impacts how much prediction errors regulate pupil dilation, we used a binned regression approach. Based on median splitting, we create two separate bins of trials with high and low contrast-difference levels for each participant. Next, we fitted the following model separately to explain the pupil dilation within each bin in a sample-by-sample fashion. Please note that this model is a reduced version of eq. (12) as we excluded contrast-difference-related regressors. All other nuisance regressors were included as control variables and normalized by z-scoring for each participant:

$$\text{Pupil} = \beta_0 + \beta_1 \cdot |\delta_t| + \beta_2 \cdot m_t + \beta_3 \cdot |u_t| + \beta_4 \cdot x_t + \beta_5 \cdot y_t \quad (13)$$

The binned regression approach results in two sets of β_1 coefficients, which reflect the extent to which absolute prediction errors impact pupil dilation in low and high contrast-difference trials.

Arousal-based analysis of behavioral residuals In order to test if and how fluctuations in pupil-linked arousal relate to belief updates beyond our behavioral model, we fit the following model to absolute single-trial belief updates:

$$\begin{aligned} |\hat{\mu}_t| = & \beta_0 + \beta_1 \cdot |p_t| + \beta_2 \cdot |\text{Pupil}| + \beta_3 \cdot |c_t| \cdot |\text{Pupil}| \\ & + \beta_4 \cdot |\text{Pupil}| \cdot |\delta_t| + \beta_5 \cdot |\text{Pupil}| \cdot |\delta_t| \cdot |c_t| \end{aligned} \quad (14)$$

Cluster-corrected permutation testing Our regression analysis, binned regression analysis and residual learning analysis resulted in coefficients estimated for each time point in the feedback time window. Thus, we used a cluster-corrected permutation-test approach for significance testing to account for multiple comparisons over time. We relied on the `ft_timelockstatistics` function by *FieldTrip* (Oostenveld et al., 2011). First, we calculated the original test statistic (t -value) for the data by comparing the conditions of interest in the case of a two-tailed test or one set of values to zero for a one-tailed test. To perform clustering, we grouped adjacent data points that exceeded a predefined threshold based on the test statistic. Next, the condition labels of the data were randomly shuffled between the two groups for the two-tailed test. Alternatively, we used a sign-flipping procedure which involves randomly changing the signs (positive or negative) of the data when comparing data from one condition to zero. The test statistic (t -value) was then calculated for the permuted dataset, and we repeated this procedure 1000 times to create a null distribution. Finally, the original test statistic was compared to this null distribution, and the p -value indicates the proportion of permuted statistics that are greater than the original statistic.

Data and code sharing

All code and processed data required to reproduce our analyses and figures are available at https://github.com/prashantig25/Perceptual_unc_aug_task_pupil.

Acknowledgements

P.G. was supported by Deutscher Akademischer Austauschdienst (DAAD) Graduate School Scholarship Programme, 2020. T.H.D. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant numbers DO 1240/6-1. R.M.C. is supported by the DFG (CI 241/3-3 and INST 272/297-1) and by a European Research Council (ERC) starting grant (ERC-2018-STG 803370). N.W.S. supported by a European Research Council Starting grant (ERCStG-REPLAY-852669), and funding from the Federal Ministry of Education and Research (BMBF) and the Free and Hanseatic City of Hamburg under the Excellence Strategy of the Federal Government and the Länder. C.F. was funded by the DFG, grant numbers FI 2309/1-1 and FI 2309/2-1. R.B. is supported by DFG grants BR 6959/2-1 and GL 984/3-1.

The authors declare no competing interests.

The authors declare no competing interests.

Supplementary material

Economic choice performance

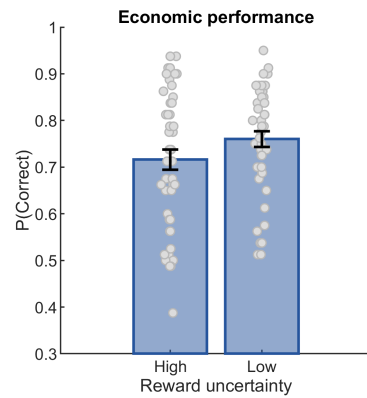


Figure S1. Choice behavior. Mean \pm standard error of the mean (SEM) economic performance, defined as the frequency of choosing the more rewarding or correct option across the two experimental conditions.

The average economic choice performance was above chance in both conditions (high reward uncertainty: mean = 0.72 ± 0.022 , $t_{45} = 9.89$, $p < 0.001$, Cohen's $d = 4.84$, low reward uncertainty: mean = 0.76 ± 0.017 , $t_{45} = 15.26$, $p < 0.001$, Cohen's $d = 6.58$).

Binned regression analysis

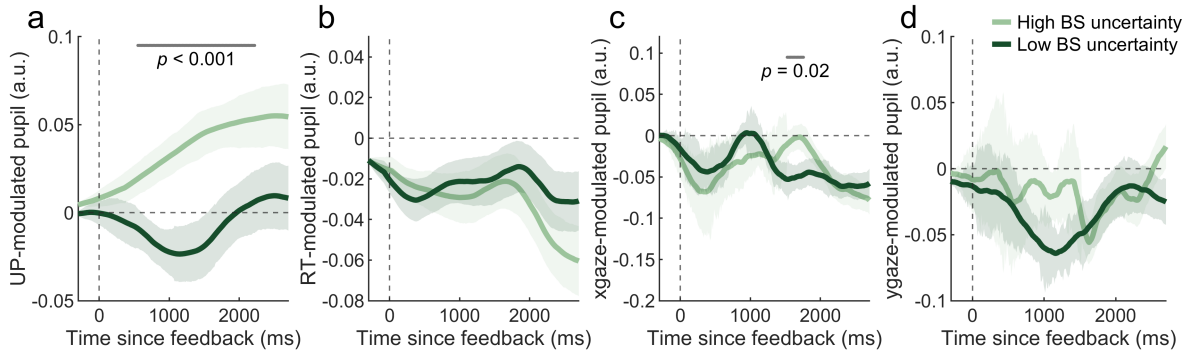


Figure S2. Binned regression approach. Regression coefficients explaining feedback-locked pupil dilation, estimated for two bins of low and high state uncertainty trials for **a**| update magnitude, **b**| reaction times, **c**| gaze position on the x-axis, **d**| gaze position on the y-axis. The gray line indicates p -values from a cluster-based permutation test of the difference between two conditions.

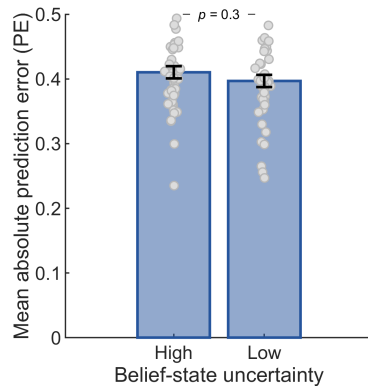


Figure S3. Absolute prediction errors for levels of state uncertainty Mean \pm standard error of the mean (SEM) absolute prediction errors for high and low levels of state uncertainty. There was no significant difference between the mean absolute prediction errors computed for trials with high state uncertainty as compared to low levels of state uncertainty.

Additional regression coefficients main analysis

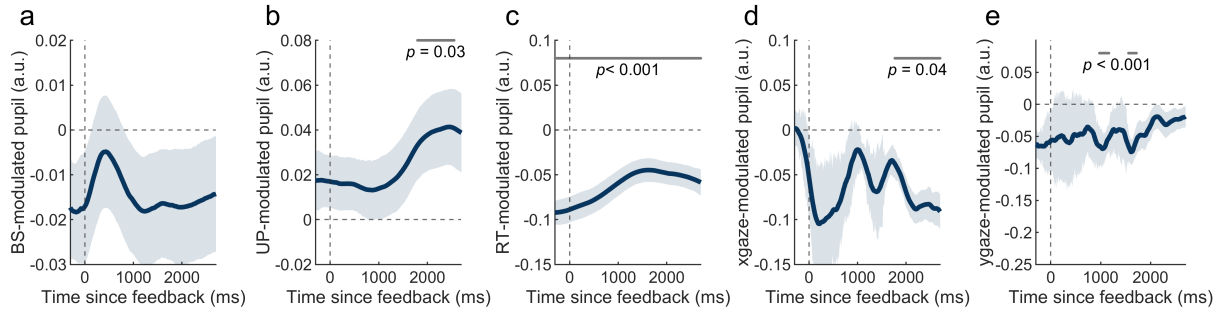


Figure S4. Influence of additional task-based and control regressors on pupil dilation post feedback. Model-estimated coefficients for each time point, explaining the impact of the main effect of **a**| contrast difference (reflecting state uncertainty), **b**| update magnitude, **c**| reaction times, **d**| gaze position on x-axis, and **e**| gaze position on y-axis on feedback-locked pupil response. The gray line indicates p -values from a cluster-based permutation test of the difference between each time course and zero.

Additional regression coefficients arousal analysis

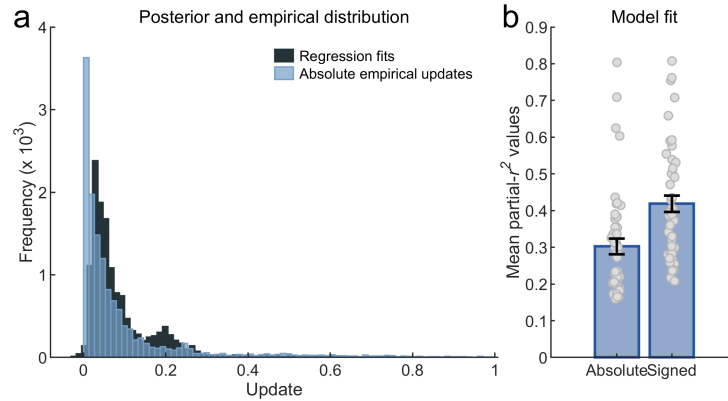


Figure S5. Model-fit assessment. **a**| A visual representation of the goodness of fit, as illustrated by the model-predicted posterior absolute updates using estimated parameters and single-trial regression data. **b**| Partial- r^2 values show the regression model was moderately effective in capturing and explaining learning data as compared to a baseline model, despite heterogeneity across participants.

To systematically compare the regression results to the empirical data, we computed the model-predicted posterior updates. Model-based updates captured the general trend in participants' learning behavior (Fig. S5a). One key difference is that empirical updates included a higher frequency of extremely small perseverative updates as compared to the model (around 0 as indicated by the blue bar in Fig. S5a). Finally, we examined the regression model's goodness of fit using partial- r^2 , which quantifies the amount of variance captured by our behavioral regression model as compared to a baseline model with no task-based predictors (Fig. S5b; see [regression diagnostics](#) for more details on the same).

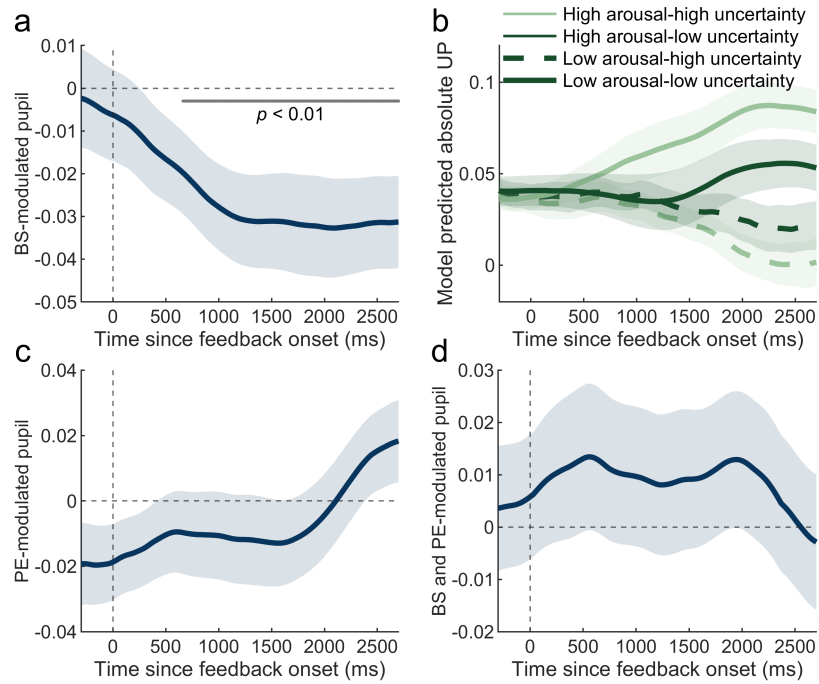


Figure S6. Contribution of arousal-related regressors in explaining belief updating beyond our behavioral model. **a** | The impact of trial-by-trial fluctuations in pupil-linked arousal to explaining residual learning was significantly modulated by the level of state uncertainty. This could suggest that pupil-linked arousal fluctuations are more crucial to explaining variance in learning behavior specifically under higher state uncertainty. **b** | Model-predicted absolute updates across bins of high and low state uncertainty. Higher state uncertainty (i.e., on average, smaller contrast differences) led to larger belief updates, for higher levels of pupil-linked arousal, as compared to trials with larger contrast differences. Moreover, the impact of trial-by-trial fluctuations in pupil-linked arousal to explaining residual learning was not significantly modulated by **c** | absolute prediction errors and **d** | and its interaction with the level of state uncertainty. The gray line indicates p -values from a cluster-based permutation test of the difference between each time course and zero.

Control regression analysis

Our main analysis approach based on eq. (12) applies the regression model to baseline-corrected pupil dilation post-feedback onset. However, to ensure that our results hold across different regression approaches, we applied two alternative control models. In the first model, pupil baseline was a regressor in the model and not removed from the signal before the analysis. In the second model, we additionally removed reaction-time components before applying the regression model. In both cases, the results were similar to our main analysis.

Baseline as a regressor

To assess the phasic pupil response independent of baseline arousal level (Gilzenrat et al., 2010; He et al., 2020; Murphy et al., 2011), we applied the following control regression model to single-trial, feedback-related and non-baseline corrected pupil diameter for each time point.

$$\begin{aligned} \text{Pupil} = & \beta_0 + \beta_1 \cdot |\delta_t| + \beta_2 \cdot |\delta_t| \cdot |c_t| + \beta_3 \cdot |c_t| + \beta_4 \cdot \text{baseline}_t \\ & + \beta_5 \cdot m_t + \beta_6 \cdot |u_t| + \beta_7 \cdot x_t + \beta_8 \cdot y_t \end{aligned} \quad (15)$$

Since the model is being fit to a non-baseline corrected measure of pupil-linked arousal, we added the mean of the pupil diameter during the baseline period (i.e., 300 ms before feedback onset) as a regressor. Thus, β_4 accounts for the impact of baseline arousal on the pupil diameter. The beta coefficients estimated from our control model aligned with the main findings. The positive main effect of absolute prediction errors and the interaction effect between prediction errors and state uncertainty qualitatively resemble the effects from our main analysis (Fig. 4a and b) and are significantly different from zero (see PE-modulated pupil, $p = 0.002$ in Fig. S5b and c).

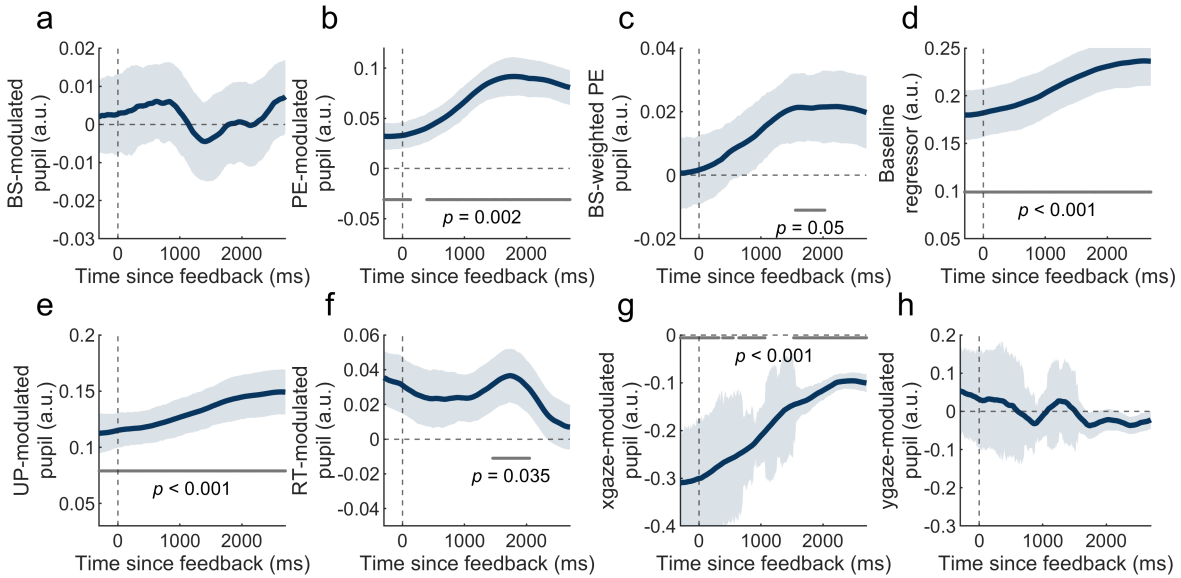


Figure S7. Control regression model with baseline as regressor. We applied a control regression model to explain non-baseline corrected pupil response at each time point in the feedback phase. To account for the impact of baseline arousal on pupil response, we added the main effect of baseline response as a regressor in the model. Model-estimated parameters for **a**| the level of state uncertainty, **b**| absolute prediction errors, **c**| interaction between state uncertainty and prediction errors, **d**| baseline pupil response, **e**| absolute updates, **f**| reaction-times, **g**| x-gaze position and **h**| y-gaze position.

Removing reaction-time components prior to analysis

We also controlled for trial-by-trial variations in the amplitude of the pupil responses caused by differences in reaction times across trials (Hoeks & Ellenbroek, 1993; Urai et al., 2017). In particular, before estimating the parameters of the regression model, we regressed out components explained by reaction time.

$$\text{Pupil}' = \text{Pupil} - (\text{Pupil}^T \cdot m_t) \cdot m_t \quad (16)$$

where "Pupil" is the original vector of pupil responses, m_t is the vector of the corresponding single-trial reaction times (log-transformed and normalized to a unit vector), and T denotes matrix transpose. We then fitted our regression model from eq. (12) to the residuals (Pupil'), which reflected pupil responses after removing components explained by single-trial reaction times. Please note that although eq. (16) accounts for the impact of reaction time on the response amplitude, we also accounted for single-trial reaction times within the regression as before.

The beta coefficients estimated from this control model were again in alignment with our main findings. The positive main effect of absolute prediction errors and the interaction effect between prediction errors and state uncertainty are comparable to the effects from our main analysis (Fig. 4a and b) and are significantly different from zero (see PE-modulated pupil, $p < 0.001$ and BS-weighted PE, $p = 0.02$ in Fig. S6b and c).

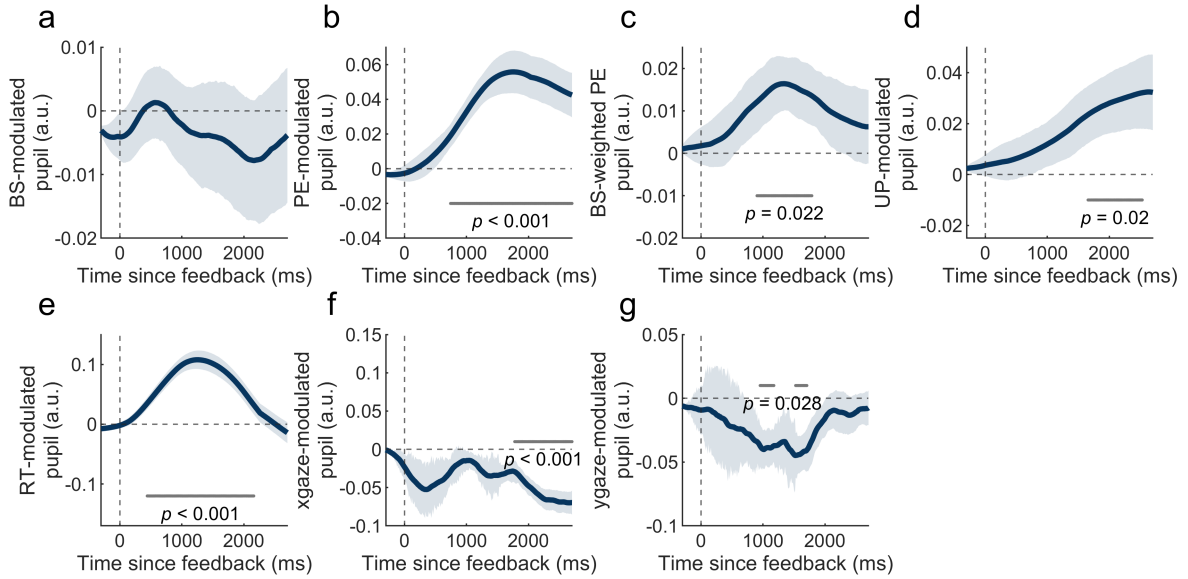


Figure S8. Control model with RT regressed. We applied a control regression model to baseline-corrected pupil response after removing the impact of reaction times at each time point in the feedback phase. Model-estimated parameters for **a**| the level of state uncertainty, **b**| absolute prediction errors, **c**| interaction between state uncertainty and prediction errors, **d**| absolute updates, **e**| reaction times, **f**| x-gaze position and **g**| y-gaze position.

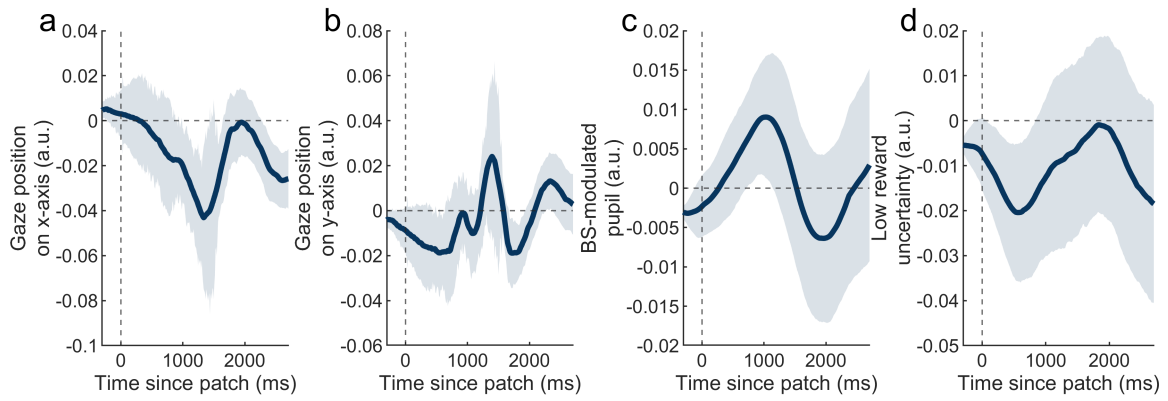


Figure S9. Influence of task-based and control regressors on pupil dilation post patch. We applied a regression model to explain baseline-corrected pupil response at each time point in the patch phase. Model-estimated parameters for **a**| x-gaze position, **b**| y-gaze position, **c**| contrast differences (reflecting state uncertainty), and **d**| low reward uncertainty.

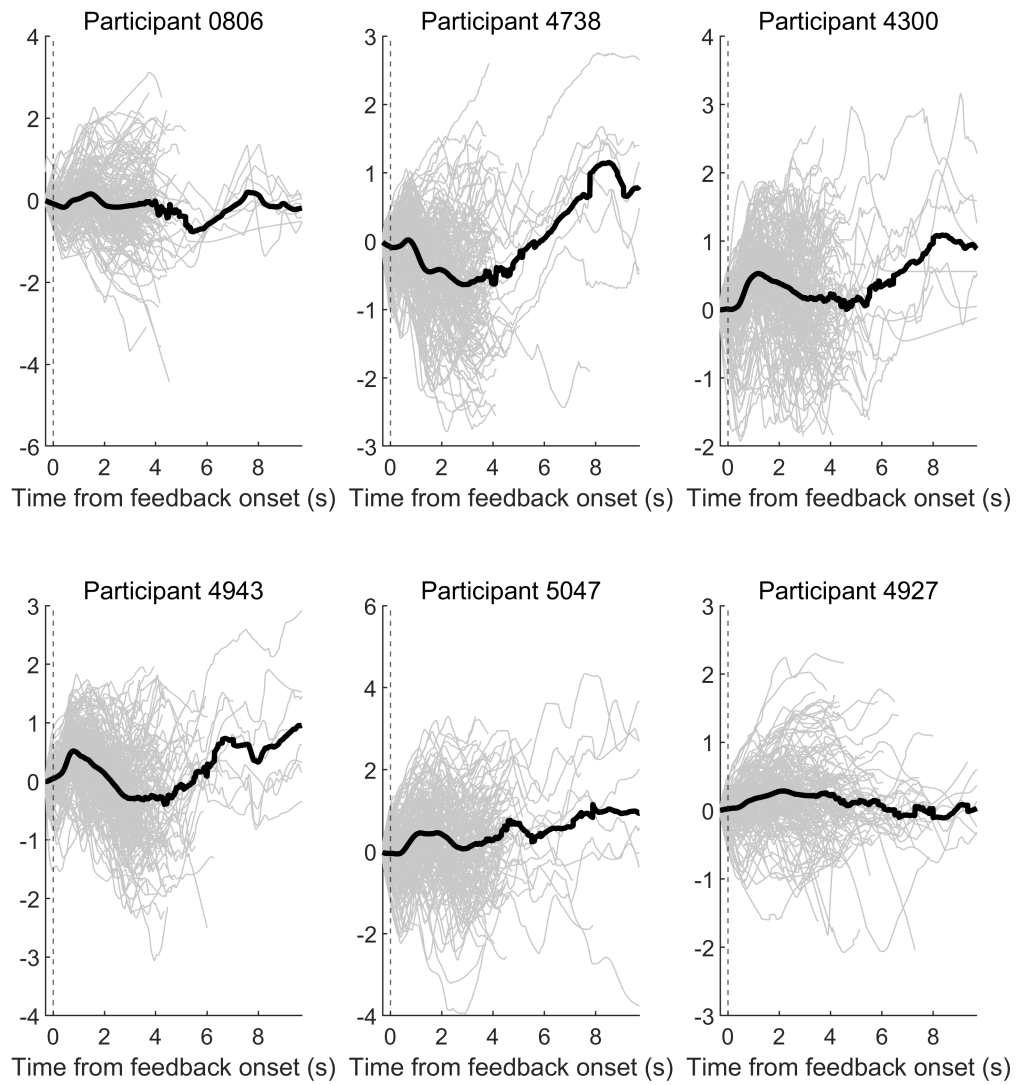


Figure S10. Single-subject average (black curve) and single-trial (gray curves) pupil response after feedback onset

References

- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450. <https://doi.org/10.1146/annurev.neuro.28.061604.135709>
- Babayan, B. M., Uchida, N., & Gershman, S. J. (2018). Belief state representation in the dopamine system. *Nature Communications*, *9*(1), 1891. <https://doi.org/10.1038/s41467-018-04397-0>
- Bear, M. F., & Singer, W. (1986). Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature*, *320*(6058), 172–176. <https://doi.org/10.1038/320172a0>
- Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, *42*(1), 33–84. [https://doi.org/10.1016/S0165-0173\(03\)00143-7](https://doi.org/10.1016/S0165-0173(03)00143-7)
- Browning, M., Behrens, T. E., Jocham, G., O’Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. <https://doi.org/10.1038/nn.3961>
- Bruckner, R., Heekeren, H. R., & Ostwald, D. (2020). Belief states and categorical-choice biases determine reward-based learning under perceptual uncertainty. *bioRxiv*. <https://doi.org/10.1101/2020.09.18.303495>
- Cazettes, F., Reato, D., Morais, J. P., Renart, A., & Mainen, Z. F. (2021). Phasic activation of dorsal raphe serotonergic neurons increases pupil size. *Current Biology*, *31*(1), 192–197.e4. <https://doi.org/10.1016/j.cub.2020.09.090>
- Chrisman, L. (1992). Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. *AAAI Press*, 183–188.
- Colizoli, O., De Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. *Scientific Reports*, *8*(1), 13702. <https://doi.org/10.1038/s41598-018-31985-3>
- Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130478. <https://doi.org/10.1098/rstb.2013.0478>
- Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 429–453. <https://doi.org/10.3758/CABN.8.4.429>
- Dayan, P., & Yu, A. J. (2006). Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network: Computation in Neural Systems*, *17*(4), 335–350. <https://doi.org/10.1080/09548980601004024>
- De Berker, A. O., Rutledge, R. B., Mathys, C., Marshall, L., Cross, G. F., Dolan, R. J., & Bestmann, S. (2016). Computations of uncertainty mediate acute stress responses in humans. *Nature Communications*, *7*(1), 10996. <https://doi.org/10.1038/ncomms10996>
- De Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T., Nieuwenhuis, S., & Donner, T. H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. *eLife*, *6*, e23232. <https://doi.org/10.7554/eLife.23232>
- De Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences*, *111*(5). <https://doi.org/10.1073/pnas.1317557111>
- Drevet, J., Drugowitsch, J., & Wyart, V. (2022). Efficient stabilization of imprecise statistical inference through conditional belief updating. *Nature Human Behaviour*, *6*(12), 1691–1704. <https://doi.org/10.1038/s41562-022-01445-0>

- Ez-zizi, A., Farrell, S., Leslie, D., Malhotra, G., & Ludwig, C. J. (2023). Reinforcement learning under uncertainty: Expected versus unexpected uncertainty and state versus reward uncertainty. *Computational Brain & Behavior*, *6*(4), 626–650. <https://doi.org/10.1007/s42113-022-00165-y>
- Ganesh, P., Cichy, R. M., Schuck, N. W., Finke, C., & Bruckner, R. (2024). Adaptive integration of perceptual and reward information in an uncertain world. *eLife*, *13*. <https://doi.org/10.7554/eLife.99266.1>
- Gershman, S. J., & Uchida, N. (2019). Believing in dopamine. *Nature Reviews Neuroscience*, *20*(11), 703–714. <https://doi.org/10.1038/s41583-019-0220-7>
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252–269. <https://doi.org/10.3758/CABN.10.2.252>
- He, M., Heindel, W. C., Nassar, M. R., Siefert, E. M., & Festa, E. K. (2020). Age-related changes in the functional integrity of the phasic alerting system: A pupillometric investigation. *Neurobiology of Aging*, *91*, 136–147. <https://doi.org/10.1016/j.neurobiolaging.2020.02.025>
- Hoeks, B., & Ellenbroek, B. (1993). A neural basis for a quantitative pupillary model. *Journal of Psychophysiology*, *7*, 315–324.
- Joshi, S., & Gold, J. I. (2020). Pupil size as a window on neural substrates of cognition. *Trends in Cognitive Sciences*, *24*(6), 466–480. <https://doi.org/10.1016/j.tics.2020.03.005>
- Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221–234. <https://doi.org/10.1016/j.neuron.2015.11.028>
- Knapen, T., De Gee, J. W., Brascamp, J., Nuiten, S., Hoppenbrouwers, S., & Theeuwes, J. (2016). Cognitive and ocular factors jointly determine pupil responses under equiluminance (T. Verguts, Ed.). *PLoS ONE*, *11*(5), e0155574. <https://doi.org/10.1371/journal.pone.0155574>
- Krishnamurthy, K., Nassar, M. R., Sarode, S., & Gold, J. I. (2017). Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nature Human Behaviour*, *1*(6), 0107. <https://doi.org/10.1038/s41562-017-0107>
- Lak, A., Nomoto, K., Keramati, M., Sakagami, M., & Kepecs, A. (2017). Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. *Current Biology*, *27*(6), 821–832. <https://doi.org/10.1016/j.cub.2017.02.026>
- Lak, A., Okun, M., Moss, M. M., Gurnani, H., Farrell, K., Wells, M. J., Reddy, C. B., Kepecs, A., Harris, K. D., & Carandini, M. (2020). Dopaminergic and prefrontal basis of learning from sensory confidence and reward value. *Neuron*, *105*(4), 700–711. <https://doi.org/10.1016/j.neuron.2019.11.018>
- Larsen, T., Leslie, D. S., Collins, E. J., & Bogacz, R. (2010). Posterior weighted reinforcement learning with state uncertainty. *Neural Computation*, *22*(5), 1149–1179. <https://doi.org/10.1162/neco.2010.01-09-948>
- Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, *37*(1), 205–220. <https://doi.org/10.1146/annurev-neuro-071013-014017>
- Manohar, S. G., & Husain, M. (2015). Reduced pupillary reward sensitivity in parkinson's disease. *npj Parkinson's Disease*, *1*(1), 15026. <https://doi.org/10.1038/npjparkd.2015.26>
- Marzo, A., Bai, J., & Otani, S. (2009). Neuroplasticity regulation by noradrenaline in mammalian brain. *Current Neuropharmacology*, *7*(4), 286–295. <https://doi.org/10.2174/157015909790031193>

- McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zaghera, E., Cadwell, C. R., Tolias, A. S., Cardin, J. A., & McCormick, D. A. (2015). Waking state: Rapid variations modulate neural and behavioral responses. *Neuron*, *87*(6), 1143–1161. <https://doi.org/10.1016/j.neuron.2015.09.012>
- Megemont, M., McBurney-Lin, J., & Yang, H. (2022). Pupil diameter is not an accurate real-time readout of locus coeruleus activity. *eLife*, *11*, 70510. <https://doi.org/10.7554/eLife.70510>
- Murphy, P. R., O’Connell, R. G., O’Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140–4154. <https://doi.org/10.1002/hbm.22466>
- Murphy, P. R., Robertson, I. H., Balsters, J. H., & O’connell, R. G. (2011). Pupillometry and p3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532–1543. <https://doi.org/10.1111/j.1469-8986.2011.01226.x>
- Nadim, F., & Bucher, D. (2014). Neuromodulation of neurons and synapses. *Current Opinion in Neurobiology*, *29*, 48–56. <https://doi.org/10.1016/j.conb.2014.05.003>
- Nassar, M. R., Bruckner, R., & Frank, M. J. (2019). Statistical context dictates the relationship between feedback-related EEG signals and learning. *eLife*, *8*, 46975. <https://doi.org/10.7554/eLife.46975>
- Nassar, M. R., McGuire, J. T., Ritz, H., & Kable, J. (2019). Dissociable forms of uncertainty-driven representational change across the human brain. *The Journal of Neuroscience*, *39*, 1713–18. <https://doi.org/10.1523/JNEUROSCI.1713-18.2018>
- Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. <https://doi.org/10.1038/nn.3130>
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. <https://doi.org/10.1155/2011/156869>
- O’Reilly, J. X., Schüffegen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *110*(38). <https://doi.org/10.1073/pnas.1305373110>
- Pfeffer, T., Keitel, C., Kluger, D. S., Keitel, A., Russmann, A., Thut, G., Donner, T. H., & Gross, J. (2022). Coupling of pupil- and neuronal population dynamics reveals diverse influences of arousal on cortical processing. *eLife*, *11*, e71890. <https://doi.org/10.7554/eLife.71890>
- Podvalny, E., King, L. E., & He, B. J. (2021). Spectral signature and behavioral consequence of spontaneous shifts of pupil-linked arousal in human. *eLife*, *10*, e68265. <https://doi.org/10.7554/eLife.68265>
- Preuschoff, K. (2011). Pupil dilation signals surprise: Evidence for noradrenaline’s role in decision making. *Frontiers in Neuroscience*, *5*. <https://doi.org/10.3389/fnins.2011.00115>
- Rasmussen, J. (2000). Designing to support adaptation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(22), 554–557. <https://doi.org/10.1177/154193120004402215>
- Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., & Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature Communications*, *7*(1), 13289. <https://doi.org/10.1038/ncomms13289>
- Reynolds, J. N. J., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. *Nature*, *413*(6851), 67–70. <https://doi.org/10.1038/35092560>
- Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, *15*(4), 507–521. [https://doi.org/10.1016/S0893-6080\(02\)00045-X](https://doi.org/10.1016/S0893-6080(02)00045-X)

- Sara, S. J. (1998). Learning by neurones: Role of attention, reinforcement and behaviour. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie*, *321*(2), 193–198. [https://doi.org/10.1016/S0764-4469\(97\)89821-6](https://doi.org/10.1016/S0764-4469(97)89821-6)
- Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human orbitofrontal cortex represents a cognitive map of state space. *Neuron*, *91*(6), 1402–1412. <https://doi.org/10.1016/j.neuron.2016.08.019>
- Starkweather, C. K., Babayan, B. M., Uchida, N., & Gershman, S. J. (2017). Dopamine reward prediction errors reflect hidden-state inference across time. *Nature Neuroscience*, *20*(4), 581–589. <https://doi.org/10.1038/nn.4520>
- Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: Neural and computational mechanisms. *Frontiers in Neuroscience*, *6*. <https://doi.org/10.3389/fnins.2012.00070>
- Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637. <https://doi.org/10.1038/ncomms14637>
- Van Den Brink, R. L., Pfeffer, T., & Donner, T. H. (2019). Brainstem modulation of large-scale intrinsic cortical activity correlations. *Frontiers in Human Neuroscience*, *13*, 340. <https://doi.org/10.3389/fnhum.2019.00340>
- Van Der Wel, P., & Van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. <https://doi.org/10.3758/s13423-018-1432-y>
- Van Slooten, J. C., Jahfari, S., Knapen, T., & Theeuwes, J. (2018). How pupil responses track value-based decision-making during and after reinforcement learning (W. Einhäuser, Ed.). *PLOS Computational Biology*, *14*(11), 1006632. <https://doi.org/10.1371/journal.pcbi.1006632>
- Vetencourt, J. F. M., Sale, A., Viegi, A., Baroncelli, L., De Pasquale, R., F. O'Leary, O., Castrén, E., & Maffei, L. (2008). The antidepressant fluoxetine restores plasticity in the adult visual cortex. *Science*, *320*(5874), 385–388. <https://doi.org/10.1126/science.1150516>
- Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. <https://doi.org/10.3758/BRM.42.3.671>
- Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279. <https://doi.org/10.1016/j.neuron.2013.11.005>
- Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. <https://doi.org/10.1016/j.neuron.2005.04.026>
- Yu, L. Q., Wilson, R. C., & Nassar, M. R. (2021). Adaptive learning is structure learning in time. *Neuroscience & Biobehavioral Reviews*, *128*, 270–281. <https://doi.org/10.1016/j.neubiorev.2021.06.024>