Task-evoked pupil responses reflect internal belief states

O. Colizoli[1,2], J.W. de Gee[1,2], A.E. Urai[1,2], T.H. Donner[1,2,3,*]

[1] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands
[3] Amsterdam Brain & Cognition, University of Amsterdam, Amsterdam, The Netherlands

*Tobias H. Donner
University Medical Center Hamburg-Eppendorf
Department of Neurophysiology and Pathophysiology, N43
Martinistraße 52
20246 Hamburg
t.donner@uke.de

Abstract

Decisions must often be made in the face of uncertain evidence. It has been proposed that the brain utilizes its neuromodulatory brainstem systems to broadcast uncertainty signals across the brain, thereby triggering changes in global brain state and neural plasticity. Recent animal work indicates that task-evoked responses in neuromodulatory systems signal two computational variables, at different times during perceptual decisions: decision confidence (i.e., the complement of uncertainty) before feedback and prediction errors (i.e., deviations from expected reward) after feedback. Neuromodulatory systems also regulate central arousal state. We here monitored pupil diameter, a marker of central arousal state, while human subjects performed a challenging perceptual choice task with delayed monetary reward. We quantified evoked pupil responses during decision formation and after reward feedback. During both intervals, decision difficulty and accuracy had interacting effects on pupil responses. This interaction reflected decision uncertainty prior to feedback and prediction error after feedback. Critically, the pattern of pupil responses during both intervals was in line with a model that uses the decision-maker's graded belief in choice accuracy to anticipate rewards and compute prediction errors. We conclude that pupil-linked arousal systems are modulated by internal belief states, in line with recent insights into dopamine signaling.

Introduction

Many decisions are made in the face of uncertainty about the state of the environment. A body of evidence indicates that decision-makers use internal uncertainty signals for adjusting subsequent choice behavior on the short-term (Kepecs et al., 2008; Meyniel et al., 2015; Urai et al., 2017), and deviations between expected and experienced rewards for learning (Montague et al., 2004; Schultz, 2005). The brain might utilize its diffusely-projecting neuromodulatory systems to broadcast such computational variables to circuits implementing inference and action selection (Schultz, 2002; Montague et al., 2004; Aston-Jones and Cohen, 2005; Yu and Dayan, 2005; Glimcher, 2011), where the neuromodulators might induce changes in the state of brain networks (Aston-Jones and Cohen, 2005; McGinley et al., 2015b) and enable synaptic plasticity (Bear and Singer, 1986; Roelfsema and Holtmaat, 2018).

Recent animal work indicates that dopaminergic midbrain neurons in particular signal two variables at different moments during a challenging perceptual decision (Lak et al., 2017): (i) decision confidence before feedback (i.e., the complement of uncertainty, which corresponds to reward expectation) and (ii) prediction error (i.e., the difference between expected and experienced reward) after receiving feedback. Critically, the prediction error signals depend on confidence (Lak et al., 2017), the agent's belief in the correctness of the choice made (see *Model Predictions* below, (Pouget et al., 2016)). Belief states have been ignored in traditional reinforcement learning models (Sutton and Barto, 1998), but incorporated in more recent ones (Dayan and Daw, 2008).

Neuromodulatory signals also govern the arousal state of the cerebral cortex, which is closely linked to non-luminance mediated changes in pupil diameter (Murphy et al., 2014a; McGinley et al., 2015a; Reimer et al., 2016; de Gee et al., 2017). Pupil responses evoked by decision tasks or micro-stimulation have commonly been associated with the noradrenergic locus coeruleus (Varazzani et al., 2015; Joshi et al., 2016; Reimer et al., 2016; de Gee et al., 2017; Liu et al., 2017). But recent work in humans also points to a contribution of dopaminergic midbrain nuclei (de Gee et al., 2017) as well as of decision uncertainty (Urai et al., 2017) to evoked pupil responses in perceptual choice tasks.

Here, we quantified pupil responses during an analogous perceptual choice task as used in recent monkey work on dopamine neurons (Lak et al., 2017), and we compared the pupil responses before and after reward feedback to predictions derived from alternative computational models of the internal variables encoded in the brain. Our goal was to (i) replicate the previously found scaling of pupil responses with decision uncertainty before

feedback (Urai et al., 2017) and (ii) test for the same scaling of pupil responses after feedback as observed for dopamine neurons (Lak et al., 2017).

Results

We monitored pupil diameter in 15 human participants performing an up vs. down random dot motion discrimination task, followed by delayed reward feedback (Figure 1). The random dot motion task has been widely used in the neurophysiology of perceptual decision-making (Gold and Shadlen, 2007; Siegel et al., 2011). Importantly, our version of the task entailed long and variable delays between decision formation and feedback, enabling us to obtain independent estimates of the pupil responses evoked by both of these events. We titrated the difficulty of the decision (by varying the evidence strength, or motion coherence, see Methods), so that observers performed at 70% correct in 2/3 of the trials in one condition ('Hard') and at 85% correct in 1/3 of the trials in the other condition ('Easy'). Correct vs. error feedback was presented after choice and converted into monetary reward, based on the average performance level across a block (25 trials), as follows: 100% correct yielded 10 Euros, 75% yielded 5 Euros, chance level (50% correct) yielded 0 Euros. The total reward earned (in Euros) was presented on the screen to participants at the end of each block.

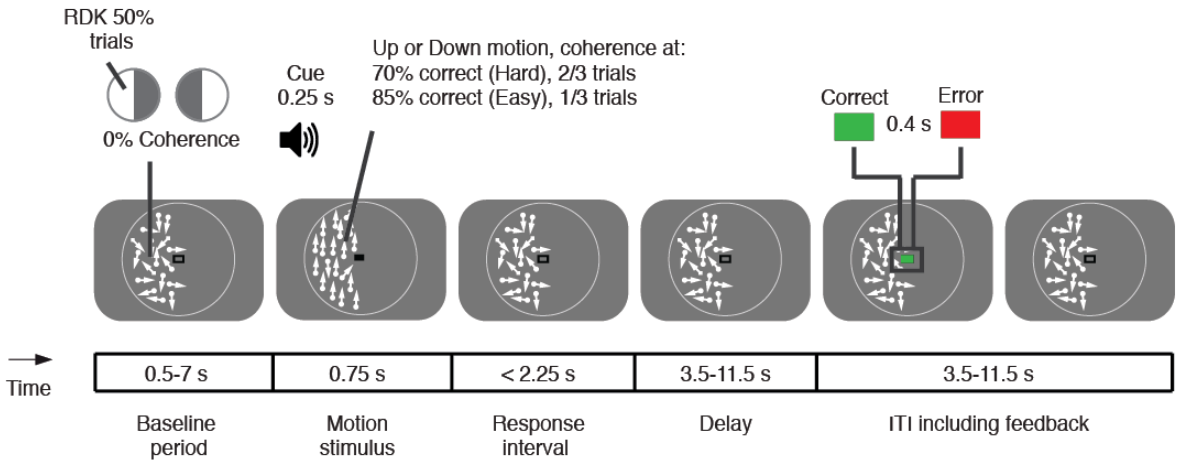


Figure 1. Perceptual choice task with delayed reward. Random dot kinematograms (RDK) were presented in one half of the visual field during each block of trials (counterbalanced). Random motion (0% coherence) was presented throughout all intervals except for the 'motion stimulus' interval, during which the RDKs to be discriminated were shown, prompted by an auditory cue (250 ms). Motion coherence of the stimulus varied from trial to trial, yielding a Hard and an Easy condition. A change from an open to a closed rectangle in the fixation region (constant luminance) prompted subjects' choice ('response interval'). After a variable delay (3.5-11.5 s) following the choice, feedback was presented that was coupled to a monetary reward (see main text). The white circle surrounding RDKs is for illustration only and was not present during the experiment.

Model predictions

We used a model based on statistical decision theory to generate qualitative predictions for the behavior of internal signals that might drive pupil-linked arousal during our task. For different intervals of the trial, these encoded either decision confidence (pre-feedback interval, (Pouget et al., 2016; Sanders et al., 2016)) or prediction errors (post-feedback interval, (Lak et al., 2017)), each as a function of correctness and stimulus difficulty (see Methods for details). The model assumed that observers categorize the motion direction based on a noisy decision variable (DV), which depended on the strength (Hard vs. Easy) and identity (Up vs. Down) of the stimulus as well as on internal noise. The model choice was governed by the difference between DV and a criterion, which was set to zero (i.e., no bias towards one over the other choice). Decision confidence, defined as the probability that a choice was correct given the evidence (Pouget et al., 2016), was a monotonic function of DV (Kepecs et al., 2008), which could be computed in the pre-feedback interval. Prediction error was computed as the difference between the probability of being correct (i.e., confidence) and the reward-linked feedback, which could be computed after observing the decision outcome (i.e., post-feedback interval). Because choice accuracy was coupled to a fixed monetary reward in our experiment (see above), confidence was the probability of obtaining the reward, in other words, reward expectation. Because the observer's internal belief state about the state of the outside world (encoded in DV) determined both reward expectation and reward prediction error we refer to this model as 'Belief State Model' in the following.

Previous work has shown that pupil responses preceding feedback and RTs followed the decision variable in the direction of uncertainty (Urai et al., 2017), which is the complement of decision confidence (with the lowest pupil dilation for the easiest, correct choices). Figure 3 shows the model predictions for decision uncertainty during the pre-feedback interval (Figure 3a) and for the complement of prediction error during the post-feedback interval (1-confidence and 1-prediction error, respectively).

The predictions of the Belief State Model contrast with those of an alternative model, in which reward expectation did not depend on the noisy internal belief (i.e., DV), but only on the strength and identity of the external stimulus (see Methods for details). This alternative model, referred to as 'Stimulus State Model' in the following, yielded predictions that matched those of traditional reinforcement learning models; by contrast, the predictions of the Belief State Model were in line with those of a reinforcement learning model based on a partially observable Markov decision process (POMDP) (Lak et al., 2017). Since the two models differed in whether confidence and prediction errors were determined by only the

external stimulus (Stimulus State Model) or by the external stimulus as well as internal noise (Belief State Model), they made distinct predictions for the scaling of pre-feedback decision uncertainty (the complement of confidence) and post-feedback prediction errors with evidence strength on correct and error trials (Figure 2, compare both rows).

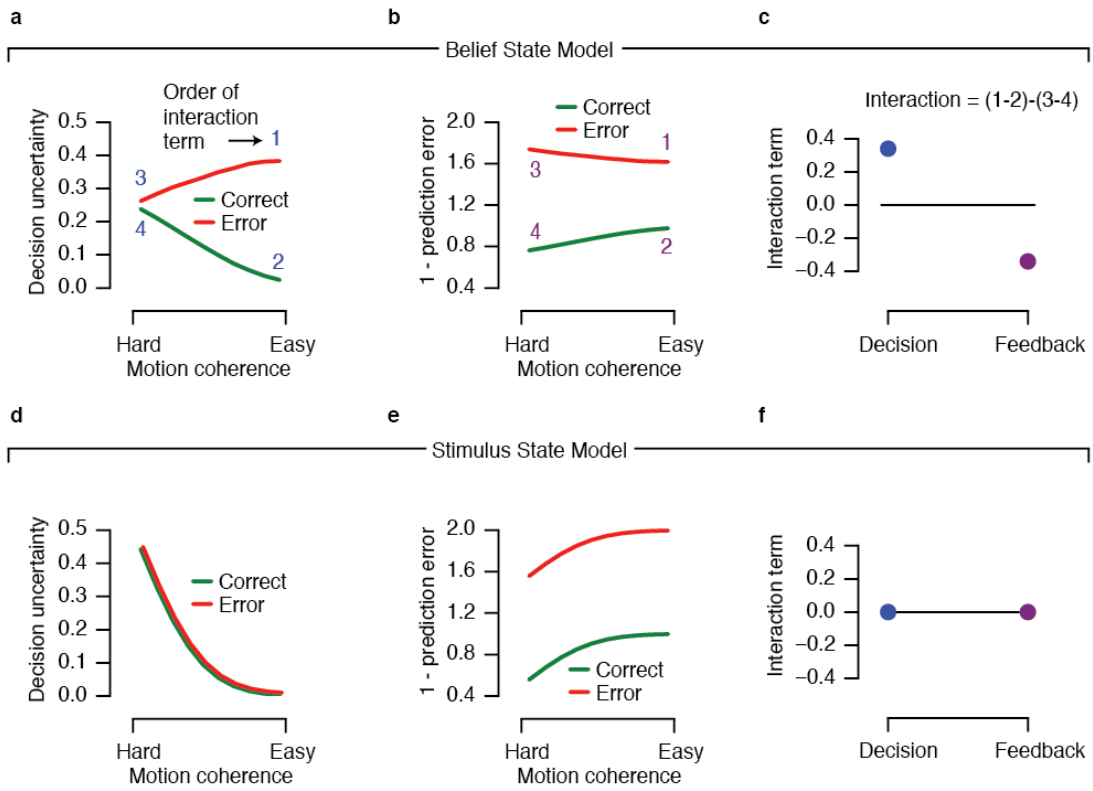


Figure 2. Alternative predictions for internal signals during pre- and post-feedback intervals of the task. (a-c) Belief State Model predictions. **(d-f)** Stimulus State Model predictions. **(a, d)** Decision uncertainty (complement of confidence) as function of motion coherence during pre-feedback interval. **(b, e)** Prediction error as function of motion coherence during post-feedback interval. **(c, f)** Interaction term computed as (Easy Error - Easy Correct) - (Hard Error - Hard Correct). See main text for model details.

Previous work on perceptual choice has shown that reaction time (RT), like pupil dilation, scales qualitatively in line with the Belief State Model predictions for decision uncertainty (Sanders et al., 2016; Urai et al., 2017; Braun et al., 2018). The same was evident in the present data: There was a main effect of accuracy, $F(1,14) = 51.57$, $p < 0.001$, and difficulty, $F(1,14) = 19.53$, $p < 0.001$, as well as an interaction effect of both, $F(1,14) = 34.95$, $p < 0.001$, on RT (see Supplementary Fig. S1, compare with Figure 2a), in line with the Belief State Model. This indicates that, in our current data, a graded, noisy DV similar to the one postulated by the Belief State Model was encoded and used for the decision process. We next tested which of the two models was more in line with the responses of

pupil-linked arousal systems. We analyzed pupil responses as a function of motion coherence and choice correctness for the two critical intervals of the trial: the phase of reward anticipation before feedback, as in previous work (Urai et al., 2017), and critically, the phase of reward prediction error signaling after feedback.

Sustained pupil response modulations during pre- and post-feedback intervals

The pupil responded in a sustained fashion during both intervals: after the onset of the coherent motion (i.e., pre-feedback) and post-feedback (Figure 3a, blue and purple lines). The pupil response remained elevated during feedback anticipation, long after decision processing (maximum of 3 s, 0.75 s stimulus duration plus response deadline of 2.25 s, see Figure 1). Upon feedback presentation, the pupil initially constricted due to the presentation of the visual feedback stimulus (see Supplementary Fig. S2) and then dilated again to a sustained level for the remainder of the post-feedback interval. Please note that we subtracted the pupil diameter during the pre-feedback period from the feedback-locked responses (see Methods), so as to specifically quantify the feedback-evoked response.

For comparison, we measured, in the same participants (separate experimental blocks), pupil responses evoked during a simple auditory detection task (button press to salient tone), which did not entail prolonged decision processing and feedback anticipation (see Methods). The resulting response, termed 'impulse response function' (IRF) for simplicity, was more transient than those measured during the main experiment: the IRF returned back to the pre-stimulus baseline level after 3 s (Figure 3a, compare IRF with the blue line). Thus, the sustained elevations of pupil diameter observed beyond that time in the main experiment reflected top-down, cognitive modulations in pupil-linked arousal due to decision processing and reward anticipation (for the responses locked to the onset of the motion stimulus), or due to reward processing (for the feedback-locked responses). We used a 'sustained' time window from 3 to 6 s (gray shaded area in Figure 3a) to extract the amplitude of these cognitive modulations of the pupil response shown in Figure 3e-r (see below).

Interacting effects of decision difficulty and accuracy on evoked pupil responses

The sustained pupil responses during both the intervals, pre- and post-feedback, scaled in line with the predictions from the Belief State Model, not the Stimulus State Model (compare Figure 3d-f with Figure 2a-c). First, pupil responses during both intervals were overall larger on error than correct trials (Figure 3b-c). The Stimulus State Model did not predict any difference between the two categories during the pre-feedback interval, because this model

was only informed by external stimulus information (RDK or feedback), not by noisy internal states. The larger pupil responses during errors in the pre-feedback interval were in line previous results, and with the idea that the central arousal level during this interval is modulated by decision uncertainty (Urai et al., 2017).

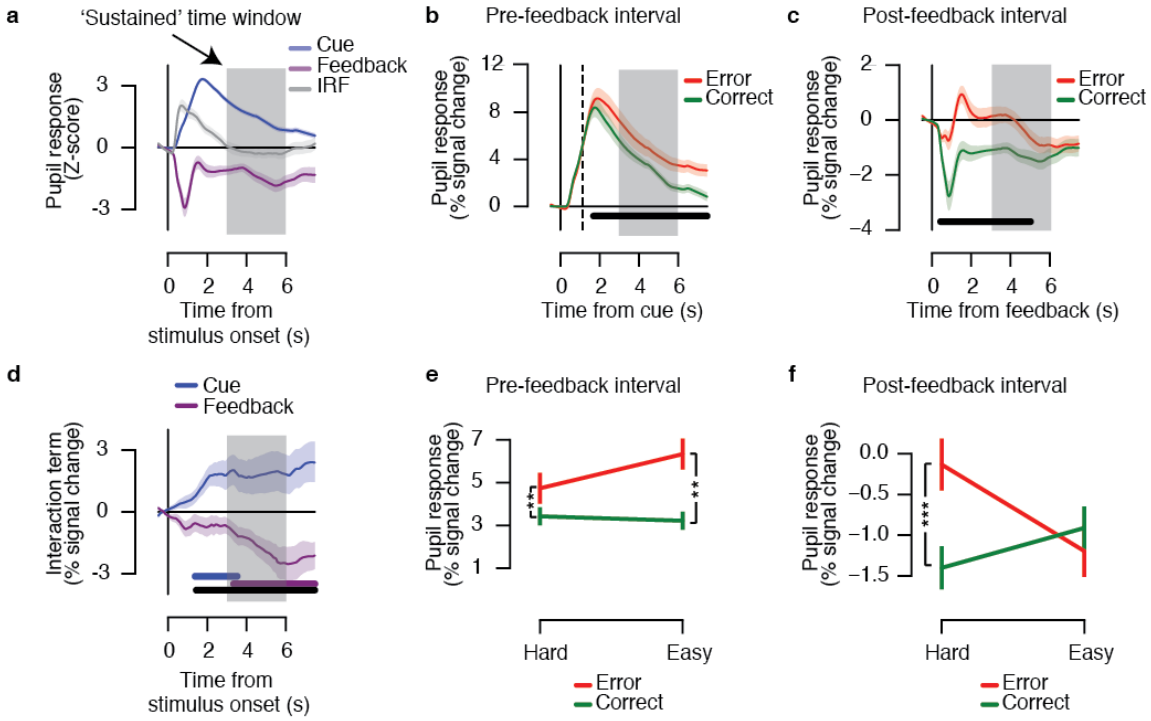


Figure 3. **Pupil responses before and after feedback reflect observers' belief state. (a)** Pupil responses locked to the motion stimulus onset cue (blue) and feedback (purple). Shown for comparison is the pupil 'impulse response' measured in the same participants when pushing a button after hearing a salient noise (IRF, see main text). Grey shading, sustained time window (3-6 s), in which IRF returned to baseline used for **(e), (f)**. **(b)** Evoked pupil responses for Correct and Error trials in pre-feedback interval. Black bar, correct vs. error, $p < 0.05$ (cluster-based permutation test). **(c)** Evoked pupil responses for Correct and Error trials in post-feedback interval. Black bar, correct vs. error, $p < 0.05$ (cluster-based permutation test). **(d)** Interaction term (Easy Error - Easy Correct) - (Hard Error - Hard Correct) for cue-locked (blue, coinciding with onset of motion stimulus) and feedback-locked (purple) responses. Horizontal bars, $p < 0.05$ (cluster-based permutation test): blue, cue-locked response against 0; purple, feedback-locked response against 0; black, difference in interaction between both responses. **(e), (f)** Mean response (in sustained time window) during both intervals, as function of difficulty and accuracy. Error bars, standard error of the mean ($N = 15$). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Second, the sustained pupil responses during both intervals exhibited a pattern of interactions between decision difficulty and accuracy as predicted by the Belief State Model but not the Stimulus State Model (compare Figure 3d to Figure 2c and Figure 2f). Hereby, the interaction was defined as (Easy Error - Easy Correct) - (Hard Error - Hard Correct). Specifically, the Belief State Model predicted a significant interaction of opposite sign for

both intervals (Figure 2c, compare blue and purple dots). That same pattern was evident in the time course of the interaction term in the pupil response: During both intervals, the interaction terms were significant, with opposite signs: positive during the pre-feedback interval and negative during the post-feedback interval (Figure 3d, blue and purple bars). Consequently, the interaction terms were significantly different from one another throughout the entire part of the sustained pupil response (Figure 3d, black bar). All three effects reached statistical significance before 3 s, which was before the shortest interval (3.5 s) to the feedback (for the pre-feedback interval) or the next trial (for the post-feedback interval) used in our experiment (Figure 1). Further, the continued interaction effects for both cue- and feedback-locked responses after 3 s were not due to contamination by the responses to the next event (feedback or next trial's cue) on trials with short delays, because the interaction terms were of opposite sign for both responses; thus, contamination would have led to cancellation of the interaction effects present in Figure 3d.

Finally, also the full pattern of sustained pupil responses for the Hard vs. Easy and Correct vs. Error conditions in both trial intervals (Figure 3e-f) resembled the pattern predicted by the Belief State Model (Figure 2a,b). In that sustained window, there was a trend towards an interaction between difficulty and accuracy during the pre-feedback interval (Figure 3e, $F(1,14) = 3.97$, $p = 0.066$; post hoc comparisons: Hard Error vs. Hard Correct, $p = 0.009$; Easy Error vs. Easy Correct, $p = 0.005$; Hard Error vs. Easy Error, $p = 0.057$; Hard Correct vs. Easy Correct, $p = 0.518$) and a significant interaction during the post-feedback phase (Figure 3f, $F(1,14) = 8.01$, $p = 0.013$; Hard Error vs. Hard Correct, $p = 0.001$; Easy Error vs. Easy Correct, $p = 0.572$; Hard Error vs. Easy Error, $p = 0.076$; Hard Correct vs. Easy Correct, $p = 0.123$).

In sum, in this perceptual choice task, sustained pupil responses during both reward anticipation (pre-feedback) as well as after reward experience (post-feedback) were qualitatively in line with the predictions from a model of reward expectation and prediction errors model, in which the computation of these internal variables depended on internal belief states.

Our current results replicate the earlier finding from a similar perceptual choice task (Urai et al., 2017), that pre-feedback pupil responses in perceptual choice scale with decision uncertainty as postulated by the Belief State Model. Our previous study focused on the pre-feedback responses and did not specifically assess the feedback-locked pupil responses (pupil measures were corrected with the same pre-trial baseline for the entire trial (Urai et al., 2017)). Further, feedback was not coupled to monetary reward in that study. We here re-analyzed the post-feedback responses in the data from that previous study in the

same way as those of the current one (see Supplementary Fig. S3). As in our current data, post-feedback responses were larger after incorrect than correct feedback (Supplementary Fig. S3 a). However, the uncertainty-dependent scaling of post-feedback responses differed between the data from our previous study (Supplementary Fig. S3 b, c) and those from the current study (Figure 3 c, d). One possible explanation for this difference, to be tested in follow-up work, might be that the prospect of receiving performance-dependent monetary reward is in fact crucial for the recruitment of pupil-linked arousal systems by uncertainty-dependent prediction errors.

Discussion

It has long been known that the pupil dilates systematically during the performance of cognitive tasks, a phenomenon referred to as task-evoked pupil response (Hess and Polt, 1960, 1964; Beatty and Kahneman, 1966; Hakerem and Sutton, 1966; Kahneman and Beatty, 1966; Kahneman et al., 1967; Simpson and Hale, 1969; Beatty, 1982). The current study shows that pupil dilation indicates decision uncertainty and reward prediction error during perceptual decision-making. Comparisons with qualitative model predictions showed that pupil responses during outcome anticipation and after reward feedback were modulated by decision-makers' (noise-corrupted) internal belief states that also governed their choices. Thus, the brain's arousal system, and specifically the neuromodulatory brainstem centers associated with pupil responses (Varazzani et al., 2015; Joshi et al., 2016; Reimer et al., 2016; de Gee et al., 2017; Liu et al., 2017), are systematically recruited by high-level computational variables. This insight is in line with theoretical accounts of neuromodulatory function (Aston-Jones and Cohen, 2005; Yu and Dayan, 2005), as well as with recent results from dopamine neurons in monkey (Lak et al., 2017). Specifically, our results, as those from dopamine neurons (Lak et al., 2017), are consistent with a reinforcement learning model (POMDP) that incorporates graded belief states in the computation of the prediction error signals that drive learning (Dayan and Daw, 2008; Lak et al., 2017).

A number of previous studies have related non-luminance mediated pupil responses to decision-making, uncertainty, and performance monitoring (Preuschoff et al., 2011; Wessel et al., 2011; Nassar et al., 2012; O'Reilly et al., 2013; de Gee et al., 2014; Lempert et al., 2015; de Berker et al., 2016), but our current results move beyond their findings in important ways. First, with the exception of (Urai et al., 2017), previous studies linking uncertainty to pupil dynamics (Preuschoff et al., 2011; Nassar et al., 2012; O'Reilly et al., 2013; de Berker et al., 2016) have used tasks in which uncertainty originated from the agent's environment. By contrast, in our task, decision uncertainty largely depended on the

observers' internal 'noise'. Importantly, the internal noise led to the qualitatively distinct predictions of both computational models for the internal variables decision uncertainty and reward prediction error (Figure 2), which drove arousal during both intervals of our task (Figure 3). Second, our earlier work, which we replicated here, also found decision uncertainty to be reflected in pupil responses prior to feedback (Urai et al., 2017). Critically, however, we here discovered that, during perceptual decision-making in the context of monetary reward, post-feedback pupil dilation reflects belief-modulated prediction error signals. While an observers' uncertainty about the accuracy of their choice can also be read out from behavioral markers such as RT (Sanders et al., 2016; Urai et al., 2017; Braun et al., 2018), no overt behavioral response is available to infer the internal variables instantiated in the brain in response to feedback experience. Thus, the insight that the post-feedback pupil dilation reports a key signal driving learning in the face of state uncertainty (Dayan and Daw, 2008) paves the way for future studies using this autonomous marker for tracking such learning signals in the brain.

Previous work on neuromodulatory systems as well as non-luminance mediated pupil dynamics has commonly used the dichotomy of (i) slow variations in baseline arousal state and (ii) rapid (so-called 'phasic') evoked responses (Aston-Jones and Cohen, 2005; de Gee et al., 2014; Murphy et al., 2014b; McGinley et al., 2015b). Our current results indicate that this dichotomy is oversimplified, by only referring to the extreme points on a natural continuum of arousal dynamics during active behavior. Our results show that uncertainty around the time of decision formation as well as the subsequent reward experience both boost pupil-linked arousal levels in a sustained fashion: Pupils remain dilated for much longer than what would be expected from a brief arousal transient (Figure 3, compare all time courses with IRF). Even in our comparably slow experimental design, these sustained dilations lasted until long after the next experimental event. This implies that the sustained evoked arousal component we characterized here contributes significantly to trial-to-trial variations in baseline pupil diameter, which have commonly been treated as 'spontaneous' fluctuations. Similarly, neuromodulatory neurons in the brainstem, also exhibit dynamics on multiple timescales (Fiorillo et al., 2003; Schultz, 2007).

Recent measurements in rodents, monkeys, and humans have shown that rapid pupil dilations reflect responses of the noradrenergic locus coeruleus (Joshi et al., 2016; Reimer et al., 2016; de Gee et al., 2017). However, in all studies, activity in other brainstem systems was also associated with pupil dilations. The behavior of pupil dilations measured in the current experiment matched the functional characteristics of dopamine neurons (Lak et al., 2017) remarkably closely, albeit with opposite sign. It is tempting to speculate that task-

evoked pupil responses also track, at least indirectly, the activity of the dopaminergic system. Another alternative is that other brainstem systems driving pupil dilations, such as the basal forebrain or superior colliculus (Joshi et al., 2016; Reimer et al., 2016; de Gee et al., 2017), exhibit the same belief-state modulated prediction error signals as the dopamine neurons characterized by (Lak et al., 2017).

Several other lines of evidence also point to an association between dopaminergic activity and non-luminance mediated pupil dilations. First, the locus coeruleus and dopaminergic midbrain nuclei are (directly and indirectly) interconnected (Weinshenker and Schroeder, 2007; Sara, 2009). In particular, both receive top-down input from the same prefrontal cortical regions (Sara, 2009), which might endow them with information about high-level computational variables such as belief states. Second, task-evoked fMRI responses of the locus coeruleus and substantia nigra are functionally coupled, even after accounting for correlations with other brainstem nuclei ((de Gee et al., 2017), their Figure 8G). Third, both neuromodulatory systems are implicated in reward processing (Weinshenker and Schroeder, 2007; Varazzani et al., 2015). Fourth, rewards exhibit smaller effects on pupil dilation in individuals with Parkinson's disease than in age-matched controls, a difference that can be modulated by dopaminergic agonists (Manohar and Husain, 2015). Finally, task-evoked pupil responses during decision formation correlate with fMRI responses in dopaminergic nuclei, again after accounting for correlations with other brainstem nuclei ((de Gee et al., 2017), their Figure 8H). Future invasive studies should establish this putative link between pupil diameter and the dopamine system.

In sum, our current study shows that internal belief states are an important factor modulating evoked responses of pupil-linked arousal systems during perceptual decision-making, during both feedback anticipation as well as after reward feedback. Our results are in line with new reinforcement learning models incorporating such belief states, as well as with the responses of dopamine neurons measured in monkey during an analogous task. Our results thus establish that pupil diameter is a non-invasive readout of an important learning signal in the brain that can be exploited by future animal and human studies. Finally, our results also add to the emerging idea that task-evoked responses in pupil diameter track responses not only of the noradrenergic system, but also of the dopaminergic system.

Methods

An independent analysis of these data for the predictive power of pupil dilation locked to motor response, for perceptual sensitivity and decision criterion has been published

previously (de Gee et al., 2017). The analyses presented in the current paper are conceptually and methodologically distinct, in that they focus on the relationship between Belief State Model predictions and pupil dilation, in particular locked to presentation of reward feedback.

Participants

Fifteen healthy subjects with normal or corrected-to-normal vision participated in the study (6 women, aged 27 ± 4 years, range 23-37). The experiment was approved by the Ethical Committee of the Department of Psychology at the University of Amsterdam. All subjects gave written informed consent. Two subjects were authors. Subjects were financially compensated with 10 Euros per hour in the behavioral lab and 15 Euros per hour for MRI scanning. In addition to this standard compensation, subjects earned money based on their task performance: 0-10 Euros linearly spaced from 50-100% accuracy per experimental session (i.e. 50% correct = 0 Euros, 75% = 5 Euros, 100% = 10 Euros). At the end of each block of trials, subjects were informed about their average performance accuracy and corresponding monetary award. Earnings were averaged across all blocks at the end of each session.

Behavioral task and procedure

Subjects performed a two-alternative forced choice (2AFC) motion discrimination task while pupil dilation was measured (Figure 1). Motion coherence varied from trial to trial, so that observers performed at 70% correct in 2/3 of trials ('hard') and at 85% correct in 1/3 of trials ('easy'). After a variable delay (3.5-11.5 s) following the choice on each trial, we presented feedback that was coupled to a monetary reward (see 'Participants').

Each subject participated in one training session and four main experimental sessions (in the MRI scanner). During the training session, subjects' individual threshold coherence levels were determined using a psychometric function fit with 7 levels, 100 trials per level, 0-80% coherence. The training session took 1.5 hours and each experimental session lasted 2 hours. During the experimental sessions, stimuli were presented on a 31.55" MRI compatible LCD display with a spatial resolution of 1920 × 1080 pixels and a refresh rate of 120 Hz.

The individual coherence levels were validated at the beginning of each experimental session in practice blocks (during anatomical scans) by checking that the subject's average accuracy across a block corresponded to 75% correct. During experimental blocks, greater motion coherence (i.e. stronger evidence strength) resulted in higher accuracy as well as

faster responses. Subjects' accuracy was higher on easy trials (M = 88.06% correct, SD = 4.26) compared to hard trials (M = 71.15% correct, SD = 3.64), $p < 0.001$. Subjects were faster to respond on easy trials (M = 1.13 s, SD = 0.13) compared to hard trials (M = 1.22 s, SD = 0.14), $p < 0.001$.

Task instructions were to indicate the direction of coherent dot motion (upward or downward) with the corresponding button press and to continuously maintain fixation in a central region during each task block. Subjects were furthermore instructed to withhold responses until the offset of the coherent motion stimulus (indicated by a visual cue). The mapping between perceptual choice and button press (e.g., up/down to right/left hand button press) was reversed within subjects after the second session (out of four) and was counterbalanced between subjects. Subjects used the index fingers of both hands to respond.

Each trial consisted of five phases during which random motion (0% coherence) was presented, with the exception of the stimulus interval: (i) the pupil baseline period (0.5-7 s); (ii) the stimulus interval consisting of random and coherent motion for a fixed duration of 0.75 s; (iii) the response window (maximum duration was 2.25 s); (iv) the delay period preceding feedback (3.5-11.5 s, uniformly distributed across 5 levels); (v) the feedback and the inter-trial interval (ITI; 3.5-11.5 s, uniformly distributed across 5 levels). Stimulus onset coincided with a visual and auditory cue. The auditory cue was presented for 0.25 s (white noise or pure tone at 880 Hz, 50-50% of trials, randomly intermixed). The visual cue was a change in the region of fixation from an open to a closed rectangle. The return of the fixation region to an open rectangle indicated to subjects to give their response (the surface areas in pixels of the open and closed rectangles were held equal in order to assure no change in overall luminance). Feedback was presented visually (green/red for correct/error) for 50 frames (0.42 s at 120 Hz). If subjects did not respond or were too fast/slow in responding, a yellow rectangle was presented as feedback on that trial.

Each block of the task began and ended with a 12-s baseline period, consisting of a fixation region (no dots). Each block of the task had 25 trials and lasted approximately 8 minutes. Subjects performed between 23 and 24 blocks yielding a total of 575–600 trials per subject. One subject performed a total of 18 blocks (distributed over three sessions), yielding a total of 425 trials. Data from one session of two subjects (12 blocks in total) and 2 blocks of a third subject were excluded from the analyses because of poor eye-tracker data quality or technical error.

Visual stimuli

Dot motion stimuli were presented within a central annulus that was not visible to the subjects (grey background, outer diameter 16.8°, inner diameter of 2.4°). The fixation region was in the center of the annulus and consisted of a black rectangle (0.45° length). Each frame consisted of 524 white dots (0.15° in diameter) within one visual hemifield (left or right; The hemifield remained constant during a block of trials and was counterbalanced between blocks. This manipulation was specific for the MRI experiment; the two hemifields were averaged in the current analysis). The proportion of 'signal' as compared with 'noise' dots defined motion coherence levels. Signal dots were randomly selected on each frame, lasted 10 frames, and were thereafter re-plotted in random locations (reappearing on the opposite side when their motion extended outside of the annulus). Signal dots moved at 7.5°/s in one of two directions (upward or downward). Noise dots were randomly assigned to locations within the annulus on each frame. To prevent tracking of individual dots, independent motion sequences (n = 3) were interleaved (Pilly and Seitz, 2009).

Eye-tracking data acquisition and preprocessing

Pupil diameter was measured using an EyeLink 1000 Long Range Mount (SR Research, Osgoode, Ontario, Canada). Either the left or right pupil was tracked (via the mirror attached to the head coil) at 1000 Hz sample rate with an average spatial resolution of 15 to 30 min arc. The MRI681 compatible (non-ferromagnetic) eye tracker was placed outside the scanner bore. Eye position was calibrated once at the start of each scanning session.

Eye blinks and saccades were detected using the manufacturer's standard algorithms (default settings). Further preprocessing steps were carried out using custom-made Python software, which consisted of (i) linear interpolation around blinks (time window from 0.1 s before until 0.1 s after each blink), (ii) band-pass filtering (third-order Butterworth, passband: 0.01–6 Hz), (iii) removing responses to blink and saccade events using multiple linear regression (responses estimated by deconvolution) (Knapen et al., 2016), and (iv) converting to percent signal change with respect to the mean of the pupil time series per block of trials.

Quantifying pre- and post-feedback pupil responses

For each trial of the motion discrimination task, two events of interest were inspected: (i) pupil responses locked to the stimulus onset and (ii) pupil responses locked to the onset of the feedback. The mean baseline pupil diameter (the preceding 0.5 s with respect to the event) was subtracted from the evoked response for each event of interest on each trial.

We extracted the mean pupil responses within the time window (3-6 s), defined by the period during which the independently measured pupil IRF returned to baseline (at the group level). The mean baseline pupil diameter (the preceding 0.5 s with respect to the event) was subtracted from the mean response within this time window for each event of interest on each trial.

Model predictions

On each trial, a decision variable, dv_i, was drawn from a normal distribution $N(\mu, \sigma)$, where μ was the sensory evidence on the current trial and σ was the internal noise. For the sensory evidence, μ ranged from -0.5 to 0.5, corresponding to the extremes of the motion coherence presented in the main experiment (where 0 = 100% random motion and 1 = 100% coherent motion). The internal noise, σ, was estimated by fitting a probit psychometric function onto the combined data across all subjects (slope $\beta = 7.5$). The standard deviation, σ, of the dv distribution is $\frac{1}{\beta} = 0.133$. The decision bound, c, was set to 0, indicating no choice bias for any observer.

For each level of evidence strength, μ = [-0.5, 0.5] in steps of 0.01, we simulated a normal distribution of dv with $\sigma = 0.133$ with ten thousand trials. The choice on each trial corresponds to the sign of dv_i. A choice was correct when the sign of dv_i was equal to the sign of μ_i.

We simulated two models, Belief State Model and Stimulus State Model, which differed in the input into the function used to compute confidence: whether the confidence is a function of dv_i or μ_i. Confidence was defined as

$$\text{Belief State Confidence} = \frac{1}{n} \times \sum_{i=1}^{n} f(|dv_i - c|) \quad (1)$$

$$\text{Stimulus State Confidence} = \frac{1}{n} \times \sum_{i=1}^{n} f(|\mu_i - c|) \quad (2)$$

where f was the cumulative distribution function of the normal distribution, transforming the distance $|dv - c|$ or $|\mu - c|$ into the probability of a correct response, for the Belief State or Stimulus State Model, respectively

$$f(x) = \frac{1}{2}[1 + \text{erf}\,(\frac{x}{\sigma\sqrt{2}})] \quad (3)$$

Decision uncertainty was the complement of confidence

$$\text{Uncertainty} = 1 - confidence \quad (4)$$

The prediction error was defined as

$$\text{Prediction error} = 1 - (feedback - confidence) \quad (5)$$

where $feedback$ was 0 or 1. Note that we here defined the prediction error to follow the direction of uncertainty, previously shown in the pupil (Urai et al., 2017). For each trial, we

computed the binary choice, the level of decision uncertainty, the accuracy of the choice and the prediction error. For plotting, we collapsed the coherence levels across the signs of μ, as these are symmetric for the up and down motion directions.

Custom Python code used to generate the model predictions can be found here: https://github.com/colizoli/belief_state_model.

Statistical analysis

Behavioral variables and pupil responses were averaged for each condition of interest per subject (N = 15). Statistical analysis of mean differences in pupil dilation of evoked responses was done using cluster-based permutation methods (Blair and Karniski, 1993). The average responses in the sustained time window were evaluated using a two-way ANOVA with factors: difficulty (2 levels: Hard vs. Easy) and accuracy (2 levels: Correct vs. Error). All post-hoc and two-way comparisons were based on non-parametric permutation tests (two-tailed).

Control experiment 1: Individual pupil impulse response functions

In order to define a sustained component of pupil responses evoked by the events of interest during the main experiment, we independently measured subjects' pupil responses evoked by simply pushing a button upon hearing a salient cue. This enabled a principled definition of the time window of interest in which to average pupil responses based on independent data. Subjects performed one block of the pupil impulse response task at the start of each experimental session (while anatomical scans were being acquired). Pupil responses evoked by a button press following an auditory cue were measured for each subject (Hoeks and Levelt, 1993). Pupils were tracked while subjects maintained fixation at a central region consisting of a black open rectangle (0.45° length) against a grey screen. No visual stimuli changed, ensuring constant illumination within a block. An auditory white noise stimulus (0.25 s) was presented at random intervals between 2 and 6 s (drawn from a uniform distribution). Participants were instructed to press a button with their right index finger as fast as possible after each auditory stimulus. One block consisted of 25 trials and lasted 2 min. Two subjects performed three blocks, yielding a total of 75-100 trials per subject. Each subject's impulse response function (IRF) was estimated using deconvolution in order to remove effects of overlapping events due to the short delay interval between subsequent trials (Knapen et al., 2016).

Control experiment 2: Pupil responses during passive viewing of feedback signals

Pupil responses evoked by the green and red fixation regions used in the main experiment were measured in a separate control experiment (see Supplementary Fig. S2; N = 15, 5 women, aged 28.5 ± 4 years, range 23-34). Three subjects were authors, two of which participated in the main 2AFC task. No other subjects from this control experiment participated in the main 2AFC task. Pupils were tracked while subjects maintained fixation at a central region of the screen. Stimuli were identical to the main 2AFC task; dot motion consisted of only random motion (0% coherence). A trial consisted of two phases: (i) the baseline period preceding the onset of a color change (1-3 s, uniform distribution), and (ii) passive viewing of the stimuli used for feedback in the main experiment: during which the fixation region changed to either red or green (50-50% of trials, randomized) for 50 frames (0.42 s at 120 Hz). This was followed by an ITI (3-6 s, uniformly distributed). Participants were instructed that they did not need to respond, only to maintain fixation. A block consisted of 25 trials and lasted 3 min. Subjects performed eight blocks of this task in the behavioral lab, yielding 200 trials per subject.

References

Aston-Jones G, Cohen JD (2005) An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. AnnuRevNeurosci 28:403–450.

Bear MF, Singer W (1986) Modulation of visual cortical plasticity by acetylcholine and noradrenaline. Nature 320:172.

Beatty J (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol Bull 91:276–292.

Beatty J, Kahneman D (1966) Pupillary changes in two memory tasks. Psychon Sci 5:371–372.

Blair RC, Karniski W (1993) An alternative method for significance testing of waveform difference potentials. Psychophysiology 30:518–524.

Braun A, Urai AE, Donner TH (2018) Adaptive History Biases Result from Confidence-weighted Accumulation of Past Choices. J Neurosci Available at: http://www.jneurosci.org/content/early/2018/01/25/JNEUROSCI.2189-17.2017.abstract.

Dayan P, Daw ND (2008) Decision theory, reinforcement learning, and the brain. Cogn Affect Behav Neurosci 8:429–453.

de Berker AO, Rutledge RB, Mathys C, Marshall L, Cross GF, Dolan RJ, Bestmann S (2016) Computations of uncertainty mediate acute stress responses in humans. Nat Commun 7:10996.

de Gee JW, Colizoli O, Kloosterman NA, Knapen T, Nieuwenhuis S, Donner TH (2017) Dynamic modulation of decision biases by brainstem arousal systems. eLife 6.

de Gee JW, Knapen T, Donner TH (2014) Decision-related pupil dilation reflects upcoming choice and individual bias. Proc Natl Acad Sci U S A 111:E618-25.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. Science 299:1898.

Glimcher PW (2011) Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. Proc Natl Acad Sci 108:15647.

Gold JI, Shadlen MN (2007) The neural basis of decision making. AnnuRevNeurosci 30:535–574.

Hakerem G, Sutton S (1966) Pupillary Response at Visual Threshold. Nature 212:485.

Hess EH, Polt JM (1960) Pupil Size as Related to Interest Value of Visual Stimuli. Science 132:349.

Hess EH, Polt JM (1964) Pupil Size in Relation to Mental Activity during Simple Problem-Solving. Science 143:1190.

Hoeks B, Levelt WJM (1993) Pupillary dilation as a measure of attention: a quantitative system analysis. Behav Res Methods Instrum Comput 25:16–26.

Joshi S, Li Y, Kalwani RM, Gold JI (2016) Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. Neuron 89:221–234.

Kahneman D, Beatty J (1966) Pupil Diameter and Load on Memory. Science 154:1583.

Kahneman D, Beatty J, Pollack I (1967) Perceptual Deficit during a Mental Task. Science 157:218.

Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455:227.

Knapen T, de Gee JW, Brascamp J, Nuiten S, Hoppenbrouwers S, Theeuwes J (2016) Cognitive and Ocular Factors Jointly Determine Pupil Responses under Equiluminance. PLoS ONE 11:e0155574.

Lak A, Nomoto K, Keramati M, Sakagami M, Kepecs A (2017) Midbrain Dopamine Neurons Signal Belief in Choice Accuracy during a Perceptual Decision. Curr Biol 27:821–832.

Lempert KM, Chen YL, Fleming SM (2015) Relating Pupil Dilation and Metacognitive Confidence during Auditory Decision-Making. PLOS ONE 10:e0126588.

Liu Y, Rodenkirch C, Moskowitz N, Schriver B, Wang Q (2017) Dynamic Lateralization of Pupil Dilation Evoked by Locus Coeruleus Activation Results from Sympathetic, Not Parasympathetic, Contributions. Cell Rep 20:3099–3112.

Manohar SG, Husain M (2015) Reduced pupillary reward sensitivity in Parkinson's disease. NPJ Park Dis 1:15026.

McGinley MJ, David SV, McCormick DA (2015a) Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. Neuron 87:179–192.

McGinley MJ, Vinck M, Reimer J, Batista-Brito R, Zagha E, Cadwell CR, Tolias AS, Cardin JA, McCormick DA (2015b) Waking state: rapid variations modulate neural and behavioral responses. Neuron 87:1143–1161.

Meyniel F, Sigman M, Mainen ZF (2015) Confidence as Bayesian Probability: From Neural Origins to Behavior. Neuron 88:78–92.

Montague PR, Hyman SE, Cohen JD (2004) Computational roles for dopamine in behavioural control. Nature 431:760.

Murphy PR, O'Connell RG, O'Sullivan M, Robertson IH, Balsters JH (2014a) Pupil diameter covaries with BOLD activity in human locus coeruleus. Hum Brain Mapp 35:4140–4154.

Murphy PR, Vandekerckhove J, Nieuwenhuis S (2014b) Pupil-Linked Arousal Determines Variability in Perceptual Decision Making. PLOS Comput Biol 10:e1003854.

Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci 15:1040–1046.

O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS (2013) Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proc Natl Acad Sci 110:E3660–E3669.

Pilly PK, Seitz AR (2009) What a difference a parameter makes: A psychophysical comparison of random dot motion algorithms. Vision Res 49:1599–1612.

Pouget A, Drugowitsch J, Kepecs A (2016) Confidence and certainty: distinct probabilistic quantities for different goals. Nat Neurosci 19:366.

Preuschoff K, 't Hart B, Einhauser W (2011) Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. Front Neurosci 5:115.

Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, Tolias AS (2016) Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nat Commun 7:13289.

Roelfsema PR, Holtmaat A (2018) Control of synaptic plasticity in deep cortical networks. Nat Rev Neurosci 19:166.

Sanders JI, Hangya B, Kepecs A (2016) Signatures of a Statistical Computation in the Human Sense of Confidence. Neuron 90:499–506.

Sara SJ (2009) The locus coeruleus and noradrenergic modulation of cognition. Nat Rev Neurosci 10:211–223.

Schultz W (2002) Getting Formal with Dopamine and Reward. Neuron 36:241–263.

Schultz W (2005) Behavioral Theories and the Neurophysiology of Reward. Annu Rev Psychol 57:87–115.

Schultz W (2007) Multiple Dopamine Functions at Different Time Courses. Annu Rev Neurosci 30:259–288.

Siegel M, Engel AK, Donner TH (2011) Cortical network dynamics of perceptual decision-making in the human brain. Front Hum Neurosci 5.

Simpson HM, Hale SM (1969) Pupillary Changes during a Decision-Making Task. Percept Mot Skills 29:495–498.

Sutton RS, Barto AG (1998) Reinforcement Learning : An Introduction. MIT Press.

Urai AE, Braun A, Donner TH (2017) Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. Nat Commun 8:14637.

Varazzani C, San-Galli A, Gilardeau S, Bouret S (2015) Noradrenaline and Dopamine Neurons in the Reward/Effort Trade-Off: A Direct Electrophysiological Comparison in Behaving Monkeys. J Neurosci 35:7866.

Weinshenker D, Schroeder JP (2007) There and Back Again: A Tale of Norepinephrine and Drug Addiction. Neuropsychopharmacology 32:1433.

Wessel JR, Danielmeier C, Ullsperger M (2011) Error Awareness Revisited: Accumulation of Multimodal Evidence from Central and Autonomic Nervous Systems. J Cogn Neurosci 23:3021–3036.

Yu AJ, Dayan P (2005) Uncertainty, Neuromodulation, and Attention. Neuron 46:681–692.

Acknowledgments: This research was supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 604102 (Human Brain Project), and the German Research Foundation (DFG): DO 1240/3-1, DO 1240/2-1, and SFB 936/A7.

Author contributions: O.C. contributed Conceptualization, Formal analysis, Investigation, Writing - original draft, Writing - review and editing; J.W.G. contributed Conceptualization, Formal analysis, Investigation, Writing - original draft, Writing - review and editing; A.E.U. contributed Formal analysis, Writing - original draft, Writing - review and editing; T.H.D. contributed Conceptualization, Resources, Supervision, Writing - original draft, Writing - review and editing;

Competing financial interests. None.

Conflict of Interest: The authors declare no competing interests exist.

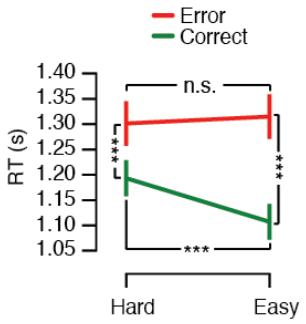
Data availability: The pupil data and model prediction code are publically available here: https://github.com/colizoli/belief_state_model. All other data and materials available upon request.

Materials & Correspondence. t.donner@uke.de (T.H.D.)

Supplementary Information: Task-evoked pupil responses reflect internal belief states

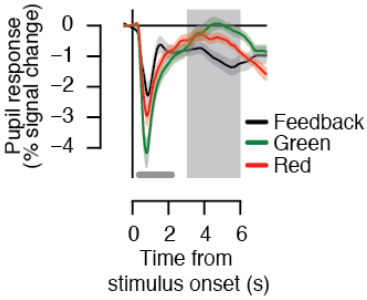
O. Colizoli[1,2], J.W. de Gee[1,2], A.E. Urai[1,2], T.H. Donner[1,2,3,*]

[1] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands
[3] Amsterdam Brain & Cognition, University of Amsterdam, Amsterdam, The Netherlands

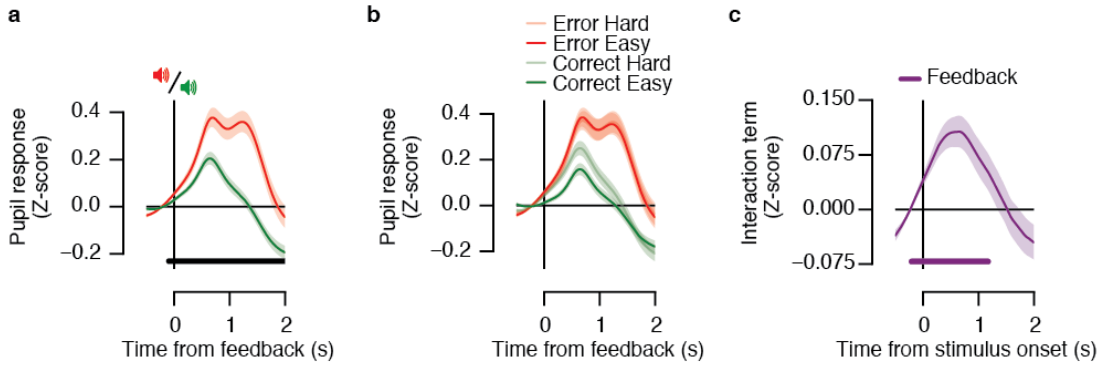


Supplementary Figure S1. RT scales with decision uncertainty.

Mean reaction times (RT) as a function of task difficulty and accuracy. Task difficulty and accuracy interacted. Error bars represent the standard error of the mean (N = 15). ***p < 0.001



Supplementary Figure S2. Pupil responses during passive viewing of feedback signals. In a control experiment (*N* = 15, 5 women, aged 28.5±4 years, range 23-34), we investigated the time course of potential differences in pupil responses evoked by red as compared with green light, regardless of whether these colors correspond to reward feedback during the perceptual choice task. Three subjects were authors, two of which participated in the main experiment. Stimuli were identical to the main 2AFC task; dot motion consisted of only random motion (0% coherence). A trial consisted of a baseline period preceding the onset of a color change (1-3 s, uniformly distributed) the red or green dot at fixation, and ITI (3-6 s, uniformly distributed). Participants passively viewed the stimuli while maintaining fixation. Pupil responses were averaged for each condition of interest per subject (*N* = 15, 200 trials per subject). The light-mediated pupil constrictions evoked by visual feedback cues during main task (grey) and in passive viewing control experiment (red, green). Horizontal bar, difference between red- and green-evoked responses, $p < 0.05$ (cluster-based permutation test, see main text). Grey shaded area, 'sustained' time window during which pupil dilation was averaged, defined by the period during which the pupil impulse response function returned to baseline and the shortest delay between events (3-6 s). The results show that (i) green and red light both evoked pupil constrictions, and (ii) green light produced slightly larger pupil constriction than red light, in an early time window (0.25-2.25 s). The difference in correct vs. error trials in pupil constriction after feedback during the main experiment continues after this early time window (see Figure 3c). Furthermore, any differences obtained *within* error and correct conditions after feedback during the main experiment cannot be explained by differences between the color-evoked responses, as the stimulus color was the same between these comparisons.



Supplementary Figure S3. Feedback-locked responses from Urai et al. (2017). Re-analysis of the data from our previously published study (Urai et al., 2017) using a similar visual perceptual choice task, however with a number of important differences specified in the following: The study used a two-interval forced choice motion coherence discrimination task; multiple levels of task difficulty were intermixed, here sorted into two categories (median split) yielding Hard and Easy conditions for comparison with the present data; delay intervals between decision and feedback, and the inter-trial-intervals were shorter than in the current study; feedback (Correct or Error) was presented by two different tones; feedback was not linked to any reward (participants' financial remuneration was not contingent on performance). **(A)** Evoked pupil responses for Correct and Error trials locked to trial-wise (auditory) feedback. Black bar, correct vs. error, $p < 0.05$ (cluster-based permutation test). Because feedback was not presented visually, there was no post-feedback pupil constriction, but dilation for all trial types. Error feedback elicited stronger dilations than correct feedback, as in the current data (compare with Figure 3c). **(B)** Pupil responses as a function of task difficulty and accuracy locked to feedback. The scaling with evidence strength was similar to pre-feedback decision uncertainty, but not to post-feedback prediction error, with smaller dilations for Correct Easy than Correct Hard responses (compare to Figure 2b). **(C)** The interaction term for task difficulty with two levels (Easy Error - Easy Correct) - (Hard Error - Hard Correct) for feedback-locked responses. Purple bar, feedback-locked response against 0, $p < 0.05$ (cluster-based permutation test). For all feedback-locked responses, the mean pupil diameter across the pre-feedback interval from -0.5 s to 0 s was subtracted from the response time courses at the single-trial level. Each condition of interest was averaged across subjects ($N = 27$).