bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Pupil dilation and P3 reflect subjective surprise about decision outcome

Jan Willem **de Gee**[1,2,3,4*], Camile M.C. **Correa**[1,5*], Matthew **Weaver**[1], Tobias H. **Donner**[1,2], Simon **van Gaal**[1]

[1]Department of Psychology and Amsterdam Brain & Cognition (ABC), University of Amsterdam, Amsterdam, the Netherlands
[2]Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[3]Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA
[4]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA
[5]Centre of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark

*Equal contribution
Correspondence: jwdegee@gmail.com (JWdG) or s.vangaal@uva.nl (SvG).

Abstract

Central to human and animal cognition is the ability to learn from feedback in order to optimize future rewards. Such a learning signal might be encoded and broadcasted by the brain's arousal systems, including the noradrenergic locus coeruleus. Pupil responses and the P3 component of event-related potentials reflect rapid changes in the arousal level of the brain. Here we ask whether and how these variables may reflect "subjective surprise": the mismatch between one's expectation about being correct and the outcome of a decision, when expectations fluctuate due to internal factors (e.g., engagement). We show that during an elementary decision-task in the face of uncertainty both physiological markers of phasic arousal reflect subjective surprise. We further show that pupil responses and P3 are unrelated to each other, and that subjective prediction error computations depend on feedback awareness. These results further advance our understanding of the role of central arousal systems in decision-making under uncertainty.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Pupil dilation at constant luminance has been related to deviant (unexpected) stimuli (Alamia et al., 2019; Bianco et al., 2020; Kamp & Donchin, 2015a; Kloosterman et al., 2015; Knapen et al., 2016; Liao et al., 2018; Murphy, Vandekerckhove, et al., 2014; Raisig et al., 2010; Van Slooten et al., 2018; Wetzel et al., 2016; Zhao et al., 2019), behavioral error awareness (Critchley, 2005; Maier et al., 2019; Murphy et al., 2016) and to key computational variables such as "uncertainty" about the state the world and/or the right course of action and subsequent "surprise" after finding out (Colizoli et al., 2018; De Berker et al., 2016; Filipowicz et al., 2020; Findling et al., 2019; Krishnamurthy et al., 2017; Lempert et al., 2015; Murphy et al., 2020; Nassar et al., 2012; O'Reilly et al., 2013; Preuschoff et al., 2011; Richer & Beatty, 1987; Urai et al., 2017; Van Slooten et al., 2018; Vincent et al., 2019). Pupil size closely tracks fluctuations in the cortical arousal state mediated by subcortical neuromodulatory systems like the noradrenergic locus coeruleus and the cholinergic basal forebrain (Aston-Jones & Cohen, 2005; Joshi & Gold, 2020; Larsen & Waters, 2018; McGinley et al., 2015). Indeed, during elementary decisions, the role of these neuromodulatory systems might be to broadcast information about momentary uncertainty and surprise (Aston-Jones & Cohen, 2005; Bouret & Sara, 2005; Doya, 2008; Lak et al., 2017; Parikh et al., 2007; Yu & Dayan, 2005), a process that can be read-out by monitoring pupil size.

In this literature, uncertainty and surprise depended lawfully on *external* factors, such as the volatility of the environment or the difficulty of the stimulus to be discriminated. Strikingly, even when we are given the same information to act on and all external factors are held constant, we will often choose differently each time when asked to make a decision (Glimcher, 2005; Gold & Shadlen, 2007; Sugrue et al., 2005; Wyart & Koechlin, 2016). Such repeated decisions tend to be associated with varying levels of uncertainty (or the inverse: confidence in being correct) (Fleming & Dolan, 2012; Fleming & Lau, 2014; Meyniel et al., 2015). Choice and confidence variability of this kind must be driven by *internal* variables.

It is unknown if pupil-linked arousal tracks subjective surprise – that is, surprise about decision outcome in the context of varying levels of decision confidence due to internal factors. This is important because deviations between objective task performance and subjective decision confidence are commonly observed, both in healthy humans as well as in several pathologies. Furthermore, it is currently unclear how peripheral markers relate to neural markers of subjective surprise. The phasic release of neuromodulators may be captured in the size of the P3 event-related potential (ERP) component, as measured with electroencephalography (EEG) over centro-parietal electrode sites (Boldt & Yeung, 2015; Brown et al., 2015; Friedman et al., 1973; Jepma et al., 2016; Kamp & Donchin, 2015; Murphy et al., 2011; Nieuwenhuis et al., 2005; Pineda et al., 1989; Polich, 2007; Steinhauer & Zubin, 1982). The P3 has been shown to scale with novelty, surprise and perceptual confidence in previous studies (Boldt & Yeung, 2015; Yeung & Summerfield, 2012). Finally, it remains an open question if and how subjective surprise also depends on the subjective awareness of the feedback stimulus. Although unconscious stimuli are known affect a plethora of cognitive processes, it is unknown how important feedback awareness is for prediction error computation (van Gaal & Lamme, 2012).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To tackle these questions, we combined an elementary perceptual decision paradigm, including explicit confidence ratings and high or low visibility feedback, with simultaneous pupil size and EEG recordings. We found (i) that both feedback-related pupil responses and P3 amplitudes reflected surprise about decision outcome, (ii) that the same pupil responses and P3 amplitudes were unrelated to each other, and (iii) that surprise about decision outcome, as reflected by the pupil and/or P3, depends on the conscious access to the feedback stimulus.

Materials and Methods

Subjects

Thirty-two students from the University of Amsterdam (23 women; aged 18-24) participated in the study for course credits or financial compensation. All subjects gave their written informed consent prior to participation, were naive to the purpose of the experiments, and had normal or corrected-to-normal vision. All procedures were executed in compliance with relevant laws and institutional guidelines and were approved by the local ethical committee of the University of Amsterdam.

Tasks

Subjects participated in three experimental sessions, separated by less than one week. We will first explain the main task, performed in session two and three, and thereafter the tasks performed in the first session. In each session, subjects were seated in a silent and dark room (dimmed light), with their head positioned on a chin rest, 60 cm in front of the computer screen. The main task was performed while measuring Pupil and EEG responses.

Main task: orientation discrimination task (sessions 2 and 3)

Stimuli were presented on a screen with a spatial resolution of 1280×720 pixels, run at a vertical refresh rate of 100 Hz. Each trial consisted of the following consecutive intervals (**Fig. 1A**): (i) the baseline interval (1.6-2.1 s); (ii) the stimulus interval (0.5 s; interrogation protocol), the start of which was signaled by a tone (0.2 s duration); (iii) the response period (terminated by the participant's response); (iv) a delay (uniformly distributed between 1.5 and 2 s); (v) the feedback interval (0.5 s), the start of which was signaled by the occurrence of a tone (0.2 s duration); (vi) a delay (uniformly distributed between 1.5 and 2 s); (vii) the feedback identity response period (terminated by the participant's response).

During Gabor presentation the luminance across all pixels was kept constant. A sinusoidal grating (1.47 cycles per degree) was presented for the entire stimulus interval. The grating was either tilted 45° (clockwise, CW) or 135° (counter-clockwise, CCW). Grating orientation was randomly selected on each trial, under the constraint that it would occur on 50% of the trials within each block of 60 trials. The grating was presented in a Gaussian annulus of 11.4 cm, with a 10.85 degrees visual angle (1.47 cycles per degree). Feedback was signaled by the Dutch word "goed" (correct feedback) or the word "fout" (incorrect feedback), from now on referred to as "correct" and "error" feedback. The words were presented for three frames just below fixation. Feedback was either masked, by presenting both forward as well as backward masks (masks1-masks2-feedback-masks3-masks4) or unmasked, by presenting only forward masks (masks1-masks2-feedback). Each mask consisted of 6 randomly scrambled letters

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(without the letters making up the words “goed” or “fout”). Masks’ types were presented two frames each. Feedback type (masked vs. unmasked) was randomly selected on each trial, under the constraint that it would occur on 50% of the trials within each block of 60 trials (**Fig. 1A**).

Throughout the main experiment, the contrast of the Gabor was fixed at the individual threshold level that yielded about 70% correct choices. Each subject’s individual threshold contrast was determined before the main experiment using an adaptive staircase procedure (Quest). The corresponding threshold contrasts yielded a mean accuracy of 70.9% correct (±0.44 % s.e.m.) in the main experiment.

Subjects performed between 12 and 17 blocks (distributed over two measurement sessions), yielding a total of 720–1020 trials per participant. Subjects were instructed to report the orientation of the Gabor, and simultaneously their decision confidence in this decision, by pressing one of four response buttons with their left or right index or middle finger: left middle finger: CCW, sure; left index finger: CCW, unsure; right index finger: CW, unsure; right middle finger: CW, sure. Subjects were also instructed to report the identity and visibility of the feedback by pressing one of four response buttons with their left or right index or middle finger: left middle finger – “error”, seen; left index finger – “error”, unseen; right index finger – “correct”, unseen; right middle finger – “correct”, seen. For analyses, we defined high visibility feedback trials as trials on which the feedback was unmasked and subjects reported it as “seen”. We defined low visibility feedback trials as trials on which feedback was masked and subjects reported it as “unseen”.

Note that indeed, the masking procedure revealed two clear categories: on unmasked trials subjects were 99.82% correct in their discrimination between feedback identity (error/correct) (s.e.m.=0.05%) whereas in the masked condition they were 71.21% correct (s.e.m.=1.69%). Note that chance level in the feedback discrimination task is not 50%, because overall Gabor discrimination performance was 70.9% and feedback presentation was veridical. Therefore, subjects may have been able to anticipate the likelihood of being correct/wrong.

Passive viewing task (session 1)

In this control task, subjects fixated their gaze at the center of the screen and passively viewed the words “goed” (correct) and “fout” (error), randomly presented for 100 times. Words were presented for three frames (100 Hz refresh rate) and were not masked.

Forced-choice visibility task (session 1)

In this control task, the words “goed” (correct) or “fout” (error) were presented in the same way as in the main experiment (see above), that is, in a masked or unmasked manner (same timings and presence or absence of masks as described above). Subjects were instructed to report the identity of the presented words, by pressing one of two response buttons with their left or right index finger: left – “error”; right – “correct” (the stimulus-response mapping was counter-balanced across trials, and was indicated on the screen after each trial). Subjects performed two blocks, yielding a total of 200 trials per participant.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In total we tested 49 subjects in the first behavioral and eye-tracking session (namely, the passive viewing and forced-choice visibility tasks). Six subjects did not enter the main experiment due to various reasons (e.g. drop-out, extensive blinking). Of the remaining 43 subjects, the 32 subjects with the lowest discrimination performance score were invited for the second and third session. Discrimination performance for the 32 included subjects varied between 49% and 73% correct. Included subjects were on average 98.87% (SEM=0.02) correct in the unmasked condition and 61.9% (SEM=0.02) correct in the masked condition. The average percentage of correct responses for masked words exceeded chance-level performance (t_{31}=11.26, $p<0.001$).

Priming task (session 1)

In this control task, subjects were instructed to respond as fast and accurately as possible to eight Dutch words, randomly selected out of five of positive (laugh, happiness, peace, love, fun) and 5 (death, murder, angry, hate, war) of negative in valence, by pressing one of two response buttons with their left or right index finger: left – negative; right – positive. Unknown to our subjects, these words were preceded by the masked words “goed” and “fout”, respectively “correct” and “incorrect”, three frames each before the positive or negative word targets (12 frames each) in 100 Hz refresh rate. This yielded congruent and incongruent trials. Subjects performed six blocks, yielding a total of 480 trials per participant.

Data acquisition

The diameter of the left eye’s pupil was tracked at 1000 Hz with an average spatial resolution of 15–30 min arc, using an EyeLink 1000 system (SR Research, Osgoode, Ontario, Canada). EEG data was recorded and sampled at 512 Hz using a BioSemi Active Two system. Sixty-four scalp electrodes were distributed across the scalp according to the 10–20 International system and applied using an elastic electrode cap (Electro-cap International Inc.) Additional electrodes were two electrodes to control for eye-movements (left eye, aligned with the pupil, vertically positioned, each referenced to their counterpart), two reference electrodes at the ear lobes to be used as reference and two electrodes for heartbeat (positioned at the left of the sternum and in the right last intercostal space).

Data analysis

Eye data preprocessing

Periods of blinks and saccades were detected using the manufacturer’s standard algorithms with default settings. The subsequent data analyses were performed using custom-made Python software. The following steps were applied to each pupil recording: (i) linear interpolation of values measured just before and after each identified blink (interpolation time window, from 150 ms before until 150 ms after blink), (ii) temporal filtering (third-order Butterworth, low-pass: 10 Hz), (iii) removal of pupil responses to blinks and to saccades, by first estimating these responses by means of deconvolution, and then removing them from the pupil time series by means of multiple linear regression (Knapen et al., 2016), and (iv) conversion to units of modulation (percent signal change) around the mean of the pupil time series from each block.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Quantification of feedback-evoked pupillary responses

We computed feedback-evoked pupillary response amplitude measures for each trial as the mean of the pupil size in the window 0.5 s to 1.5 s from feedback, minus the mean pupil size during the 0.5 s before the feedback. This time window was chosen to be centered around the peak of the pupil response to a transient event (like the feedback in our task; **Fig. 2A**) (de Gee et al., 2014; Hoeks & Levelt, 1993).

EEG data preprocessing

Standard pre-processing steps were performed in EEGLAB toolbox in Matlab. Data were bandpass filtered from 0.1 to 40 Hz off-line for ERP analyses. Epochs ranging from -1 to 2 seconds around feedback presentation were extracted. Linear baseline correction was applied to these epochs using a -200 to 0 ms window. The resulting trials were visually inspected and those containing artifacts were removed manually. Moreover, electrodes that consistently contained artifacts were interpolated, entirely or per bad epoch. Finally, using independent component analysis, artifacts caused by blinks and other events not related to brain activity were manually removed from the EEG data.

Quantification of feedback-related ERP components

We focused on ERP components related to feedback processing with different latencies and topographical distributions. To zoom in on these specific components, a central region of interest (ROI) was defined (including the averaged signal of electrodes F1, Fz, F2, FCz, FC1, FC2, Cz, C1, C2, CPz, CP1, CP2, Pz, P1, P2). The P3 window was defined from 0.5 to 0.8 seconds after feedback presentation.

Behavioral analyses and statistical comparisons

Behavioral and statistical analyses were performed in Python. We only considered unmasked trials reported as seen as high visibility feedback (99% of unmasked trials) and masked trials indicated as unseen as low visibility feedback (89.2% of masked trials). RT was defined as the time from stimulus offset until the button press. We used 2x2 repeated measures ANOVA to test for the main effect of being correct, and for the interaction effect between correctness and confidence. With a 2x2x2 repeated measures ANOVA we tested whether these main and interaction effects were different between the high and low visibility conditions. We used the paired-samples t-test to test for differences in RT, accuracy or choices between high and low confidence trials, and between congruent and incongruent priming conditions. We used Pearson correlation to quantify the relationship between pupil responses and P3 amplitudes.

Data and code sharing

The data are publicly available on [to be filled in upon publication]. Analysis scripts are publicly available on [to be filled in upon publication].

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

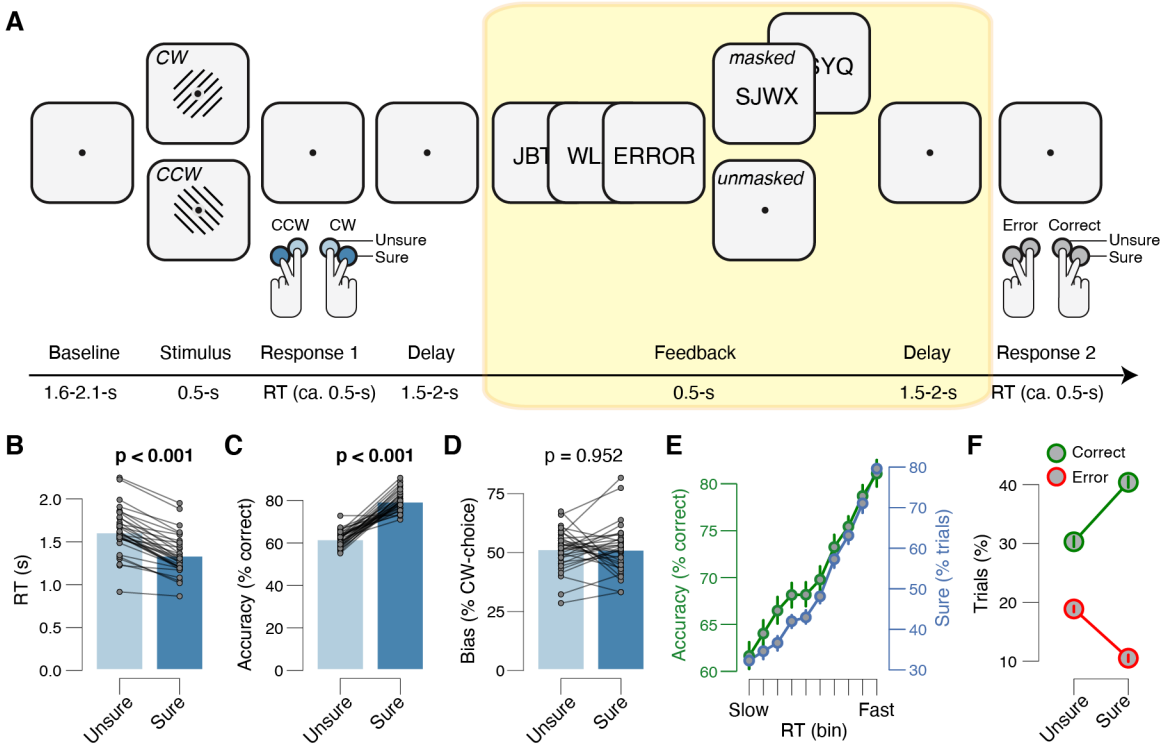

Figure 1. Task and behavior. (A) Sequence of events during a single trial. Subjects reported the direction and level of confidence in the decision about a Gabor patch by pressing one of four buttons (CCW=counter-clock-wise, CW=clock-wise; CCW sure; CCW unsure, CW unsure, CW sure). After the decision interval, veridical feedback was presented indicating the correctness of the response. Subjects reported the identity and visibility of the feedback stimulus (the word "error" or "correct" in Dutch) by pressing one of four buttons (seen error; unseen error; unseen correct; seen correct; see Methods for details). **(B)** RT separated for high/low confidence decisions. Data point, individual subjects; stats, paired-samples t-test. **(C)** As B, but for accuracy. **(D)** As B, but for choice bias. **(E)** Proportion of correct (green) and high-confident (blue) choices for 10 RT-defined bins. Error bars, s.e.m. **(F)** Proportion of trials sorted by correctness (error, correct) and confidence (high, low). Error bars, s.e.m. Panels B-F: group average (N=32).

Feedback-related pupil response and P3 amplitudes report a subjective prediction error

During simultaneous pupillometry and EEG recordings, thirty-two human subjects performed a challenging contrast orientation discrimination task (three experimental sessions per subject, on different days). On each trial this involved discriminating the orientation (clockwise [CW] vs. counter-clockwise [CCW]) of a low-contrast Gabor, explicit confidence ratings and feedback (**Fig. 1A**). The Gabor's contrast was adjusted individually such that each subject performed at about 70% correct (**Fig. 1C**; Materials and Methods). Subjects simultaneously indicated their CW/CCW-choice and the accompanying confidence in that decision (sure/unsure; type 1 confidence (Galvin et al., 2003), see **Fig. 1A**). These explicit ratings provided a window into the trial-to-trial fluctuations of decision confidence, which may shape prediction error signals after decision outcome (feedback), and physiological correlates thereof. The Dutch words for "error" or "correct" provided feedback about the correctness of the preceding CW/CCW-choice. Feedback was preceded and followed by masks (random letters) on 50% of trials. This was done to test whether uncertainty about whether the decision was correct or incorrect, because of varying levels of feedback awareness, may affect phasic measures of central arousal state. At the end of the trial, subjects had to indicate the subjective visibility and identity (the word "error" or "correct") of the feedback stimulus (**Fig. 1A**). This

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

allowed us to post-hoc sort trials based on the combination of masking strength and subjective visibility (Materials and Methods).

Subjects' choice behavior indicated that they successfully introspected perceptual performance: subjects were faster and more accurate when they were confident in their decision (**Fig. 1B,C**), a typical signature of confidence (Kamp & Donchin, 2015b; Meyniel et al., 2013). There was no relationship between confidence and decision bias (**Fig. 1D**). In line with earlier work (Sanders et al., 2016; Urai et al., 2017), reaction times (RTs) predicted accuracy and confidence, with more accurate and confident choices for faster RTs (**Fig. 1E**). Taken together, these results suggest that subjects in our task were able to introspect perceptual performance well.

Negative feedback ('error') was more surprising than positive feedback ('correct'), because subjects performed well above chance (~71% correct). Negative feedback should be especially surprising when subjects were relatively sure about the correctness of the preceding choice. In contrast, positive feedback should be least surprising when they were relatively sure about the correctness of the preceding choice. In line with this intuition, trial counts followed the expected ordering (from least to most often / from most to least surprising): sure-error, unsure-error, unsure-correct, sure-correct (**Fig. 1F**). For trial counts, there was thus a significant main effect for correctness ($F_{1,31}$=2663.43, p<0.001) and an interaction effect between correctness and confidence ($F_{1,31}$=129.23, p<0.001). Any physiological variable that encodes a subjective prediction error (surprise about decision outcome) should follow a similar pattern. More specifically, although the main effect of correctness (feedback: error vs correct) may partly reflect typically observed error monitoring processes (Cohen et al., 2011; Ullsperger et al., 2014), which are not necessarily related to prediction errors, because it can be triggered purely by the type feedback received (error or correct). It is however especially the interaction between confidence and correctness that we consider a signal of subjective prediction error computation.

Using the same 2x2 ANOVA logic, we tested whether the amplitude of the feedback-related pupil response and/or P3 component reflect a subjective prediction error. Indeed, feedback-related pupil responses (to high visibility feedback) were larger for negative versus positive feedback (main effect of correctness: $F_{1,31}$=43.33, p<0.001; **Fig. 2A,B**). This error vs. correct difference was larger when subjects' decision confidence was high vs. low (interaction correctness x confidence: $F_{1,31}$=11.30, p=0.002; **Fig. 2A,B**). We verified that the correctness main effect was not driven by any low-level stimulus characteristics, such as luminance, or the intrinsic valence of the words used as feedback (e.g. being of positive/negative valence; Materials and Methods; **Fig. 1I**). The feedback-related P3 exhibited similar functional properties as the feedback-related pupil responses (**Fig. 2D,E**): the P3 component was larger for feedback indication an error vs. correct decision ($F_{1,31}$=66.72, p<0.001) and this effect interacted with decision confidence ($F_{1,31}$=6.04, p=0.020, see headmaps in **Fig. 2D** for topographical distributions of these effects).

If the pupil- and P3-responses are driven by the same central (e.g., neuromodulatory) process, as postulated in one influential account (Sander Nieuwenhuis et al., 2005), then both measures should not only exhibit the same dependence on experimental conditions on average (**Fig. 2A-F**), but also their main and interaction effects should correlate across

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

subjects. We did not find evidence for this (**Fig. 2G,H**). This suggests that pupil size and the P3 are driven by distinct neural processes, both of which are sensitive to decision confidence and prediction errors. We used the reaction times, a sensitive measure of confidence (**Fig. 1E**), to visualize and quantify the pupil- and P3-reported prediction errors in a more fine-grained fashion. To that end, we used the same 2x2 ANOVA logic. The feedback-related pupil responses were larger for negative compared to positive feedback ($F_{1,31}$=48.77, p<0.001) and this effect interacted with RT ($F_{1,31}$=12.48, p=0.001, **Fig. 1C**). Likewise, the feedback-related P3 amplitudes were larger for negative compared to positive feedback ($F_{1,31}$=71.91, p<0.001) and this effect interacted with RT ($F_{1,31}$=6.08, p=0.019, **Fig. 1F**).

Taken together, we conclude that both physiological variables, feedback-related pupil responses and P3 amplitudes, report a subjective prediction error, when feedback is presented fully consciously.

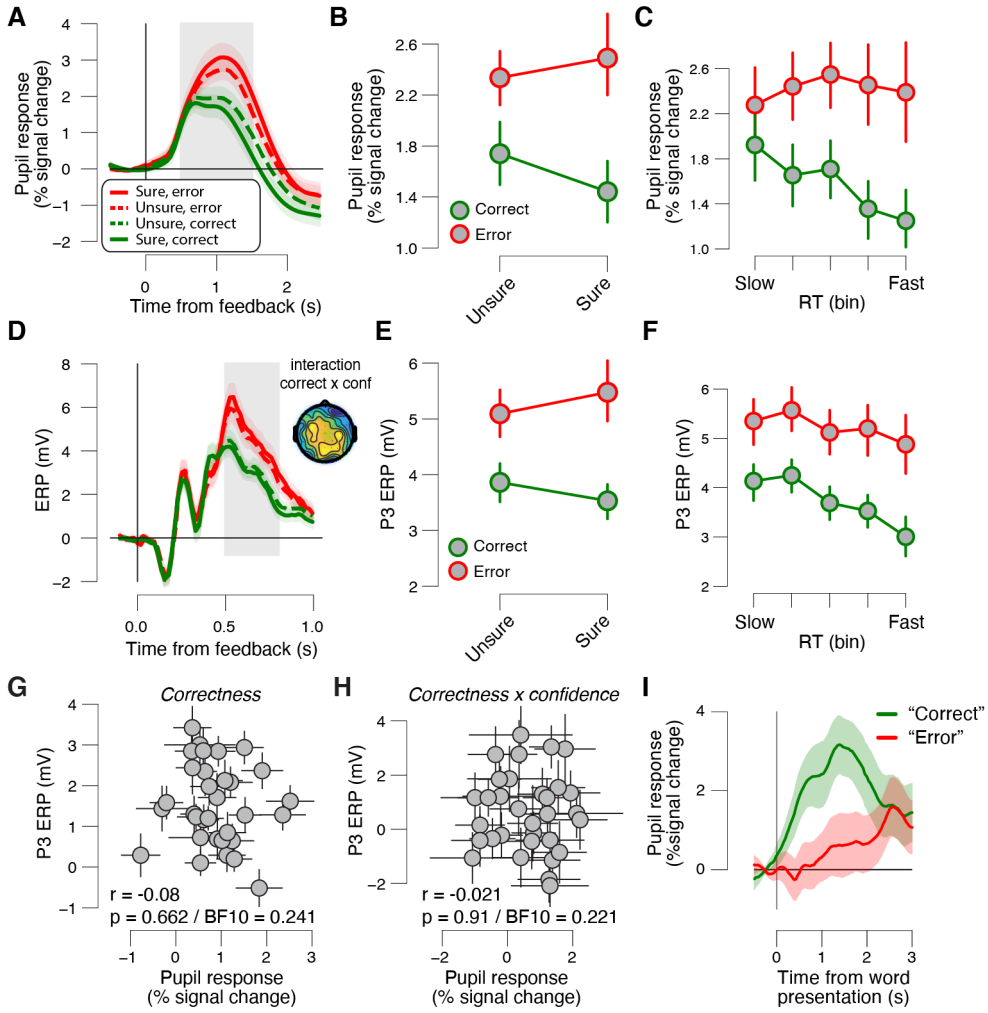

Figure 2. Feedback-related pupil response and P3 report a subjective prediction error. (A) High visibility feedback-related pupil time course, sorted by correctness (error, correct) and confidence (high, low). Grey box, interval for averaging pupil response values on single trials. **(B)** High visibility feedback-related pupil responses sorted by correctness (error, correct) and confidence (high, low). **(C)** High visibility feedback-related pupil responses sorted by correctness (error, correct) and RT. **(D)**, As A, but for the high visibility feedback event-related potential (ERP) time courses. Head map, interaction effect between correctness and confidence (map limits [-1 1]). **(E,F)**, as B,C but for high visibility feedback-

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

related P3 scalar amplitude measures. **(G)** High visibility feedback-related P3 correctness main effect plotted against pupil response correctness main effect. Stats, Pearson correlation; datapoints, individual subjects (N=32); error bars, 60% confidence intervals (bootstrap). **(H)** As G, but for correctness x confidence interaction effects. **(I)** High visibility event related pupil time course sorted by the words "correct" and "error" during a passive viewing experiment (Materials and Methods). All panels except G,H: group average (N=32); shading or error bars, s.e.m.

Physiological correlates of subjective prediction errors depend on feedback awareness

Feedback-related pupil responses and P3 amplitudes did not report a subjective prediction error after low visibility feedback (Materials and Methods). For the feedback-related pupil responses, there was no significant main effect of correctness ($F_{1,31}$=4.04, p=0.053) nor an interaction effect thereof with confidence ($F_{1,31}$=0.04, p=0.840, **Fig. 3A,B**). Likewise, for the feedback-related P3 amplitudes, there was no significant main effect of correctness ($F_{1,31}$=3.67, p=0.064) nor an interaction effect thereof with confidence ($F_{1,31}$=0.60, p=0.442, **Fig. 3D,E**). As for the high visibility feedback, the pupil- and P3-main and interaction effects were not correlated across participants (**Fig. 3G,H**).

We did also not find evidence for a prediction error using the more fine-grained framework based on RTs. For the feedback-related pupil responses, there was a barely significant correctness main effect ($F_{1,31}$=4.46, p=0.043) and no interaction effect thereof with RT ($F_{1,31}$=0.73, p=0.575, **Fig. 3C**). For feedback-related P3 amplitudes, there was no significant correctness main effect ($F_{1,31}$=3.65, p=0.065) nor an interaction effect thereof with confidence ($F_{1,31}$=0.75, p=0.555, **Fig. 3F**).

We ruled out the possibility that the low visibility feedback was too weak (because of masking) to drive a potential prediction error. A behavioral priming experiment (Materials and Methods) with the same stimuli and stimulus timings showed typical priming effects, both in regard to RT and accuracy. We observed faster responses and higher accuracy for congruent prime-target pairs versus incongruent pairs (**Fig. 3I**).

Feedback-related pupil responses and P3 amplitudes reported subjective prediction errors significantly better after high versus low visibility feedback. For the feedback-related pupil responses, there was a significant visibility x correctness interaction effect ($F_{1,31}$=14.90, p<0.001) and a visibility x correctness x confidence interaction effect ($F_{1,31}$=6.99, p=0.013). Likewise, for the feedback-related P3 amplitudes, there was a significant visibility x correctness interaction effect ($F_{1,31}$=48.65, p<0.001) but no significant visibility x correctness x confidence interaction effect ($F_{1,31}$=3.01, p=0.093).

In sum, feedback-related pupil responses and P3 amplitudes only reflected a subjective prediction error after high visibility feedback (interaction correctness x confidence). Our results indicate that full visibility of decision feedback is critical to drive a subjective prediction error response.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

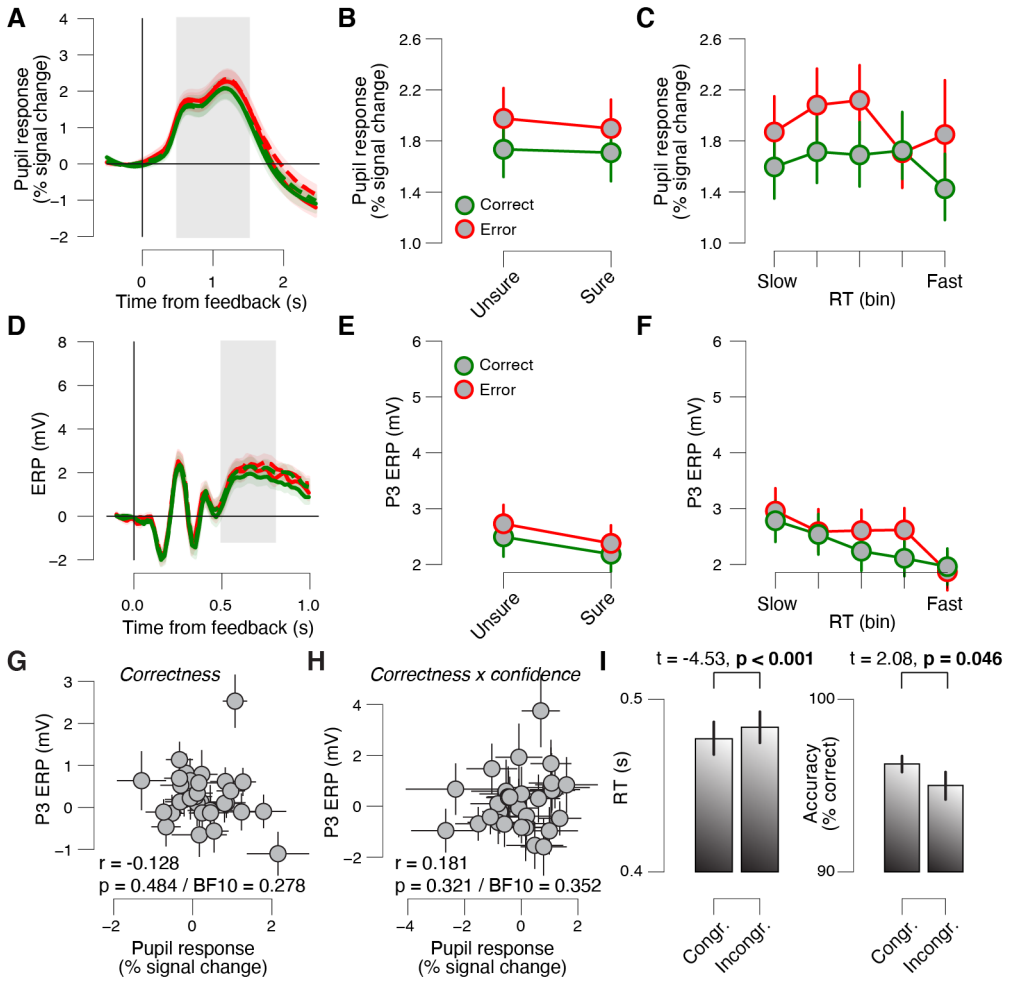

Figure 3. Physiological correlates of subjective prediction errors depend on feedback awareness. (A) High visibility feedback-related pupil time course, sorted by correctness (error, correct) and confidence (high, low). Grey box, interval for averaging pupil response values on single trials. **(B)** High visibility feedback-related pupil responses sorted by correctness (error, correct) and confidence (high, low). **(C)** High visibility feedback-related pupil responses sorted by correctness (error, correct) and RT. **(D)**, As A, but for the high visibility feedback event-related potential (ERP) time courses. Head map, interaction effect between correctness and confidence (map limits [-1 1]). **(E,F)**, as B,C but for high visibility feedback-related P3 scalar amplitude measures. **(G)** Low visibility feedback-related P3 correctness main effect plotted against pupil response correctness main effect. Stats, Pearson correlation; datapoints, individual subjects (N=32); error bars, 60% confidence intervals (bootstrap). **(H)** As G, but for correctness x confidence interaction effects. **(I)** Reaction times (left) and accuracy (right) sorted by congruency (congruent, incongruent) showing typical behavioral priming effects (Materials and Methods). All panels except G,H: group average (N=32); error bars, s.e.m.

Discussion

Here we show that two markers of phasic arousal, pupil size and the P3 ERP component, reflect subjective prediction error responses. This subjective prediction error response was only observed when the feedback stimulus, indicating the correctness of the previous decision, was presented at full visibility. These results were not driven by any low-level stimulus characteristics, such as luminance, or the intrinsic valence of the words used as feedback (e.g. being of positive/negative valence, **Fig. 2I**). The reported findings advance the current

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

knowledge about the role of decision confidence and conscious awareness in prediction error computations in several important ways.

For the first time (to our knowledge), we reveal that prediction error responses are truly defined by subjective decision confidence, despite equal task difficulty. Previous studies have revealed that pupil dilation reflects decision uncertainty and prediction error computation during perceptual choices (Colizoli et al., 2018; Joshi & Gold, 2020; Urai et al., 2017) when task difficulty was manipulated. In these studies, humans performed a random dot motion task incorporating easy and difficult trials depending on the strength of motion coherence. Pupil responses were larger for performance feedback informing that the decision was erroneous vs. correct and this effect was modulated by trial difficulty, in such a way that the pupil dilated most for erroneous decisions based on strong evidence (strong prediction error) and least for correct decisions based on strong evidence (no prediction error). However, several studies have shown that subjective reports of decision confidence do not necessarily track experimental manipulations of task difficulty, for example because confidence estimations are biased due to individual differences in sensitivity to evidence strength or affective value (Fleming & Lau, 2014; Lebreton et al., 2018; Zylberberg et al., 2014). Because we interrogated subjective decision confidence on every single trial, this allowed us to perform post-hoc trial sorting based on trial-by-trial fluctuations in confidence under equal task settings. Thereby we were able to link feedback processing directly to subjective confidence estimations, establishing direct evidence for subjective prediction error computation in the pupil and P3 ERP component.

Further, we show that full visibility of performance feedback is crucial for subjective prediction error computation. Although there is consensus that some perceptual and cognitive processes may unfold in the absence of awareness, it is highly debated which functions (if any) may need consciousness to emerge (Dehaene & Naccache, 2001; Hommel, 2007; Kunde et al., 2012; van Gaal et al., 2012). Many perceptual and cognitive processes may partly unfold unconsciously, typically demonstrated in masked priming studies, in which a task-irrelevant unconscious stimulus facilitates responding to a subsequent task-relevant conscious stimulus (Kiefer et al., 2011; Kiesel et al., 2007; Lamme, 2010). These “simple” priming effects are typically explained by assuming that the fast feedforward sweep of neural processing is relatively unaffected by masking and is able to unconsciously affect ongoing behavioral responses (Dehaene & Changeux, 2011; van Gaal et al., 2012). We also observed here that the same masked stimuli used as feedback in the main experiment (the words error/correct) could induce behavioral priming when presented in the context of a masked priming task (**Fig. 3**). However, when the same stimuli were used as feedback stimuli in a perceptual decision task, no subjective prediction error responses were observed (correctness x confidence interaction). Speculatively, error detection mechanisms (main effect of correctness) could still be observed when feedback was masked in the current task design (**Fig. 3**). Although evidence was statistically relatively weak, the main effects of correctness, especially in pupil size, were significant. This may not be overly surprising, because previous studies have shown that error detection mechanisms may unfold (at least partially) in the absence of error awareness (Charles et al., 2013; Cohen et al., 2009; Nieuwenhuis et al., 2001; Overbeek et al., 2005; Shalgi, 2012). However, our results revealed more importantly an absence of confidence x correctness interactions on low visibility feedback. This may suggest that to

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

incorporate subjective confidence in feedback-driven prediction error computations, awareness of the decision outcome (feedback) is crucial. This may suggest that subjective prediction error computation cannot rely on feedforward responses alone, in contrast to e.g. masked priming, and requires (bidirectional) interactions (i.e. recurrent processing) between higher-order and lower-order regions, a phenomenon mainly observed when stimuli are presented above the threshold of conscious perception (Dehaene & Changeux, 2011; van Gaal & Lamme, 2012). Further unraveling the underlying neural processes dissociating "objective error processing" from "subjective prediction error" computation is important for further understanding the potential scope and limits of unconscious information processing.

Previous studies have shown that pupil size is sensitive to implicit surprise and effort invested in a cognitive task. For example, it has recently been demonstrated that pupil size increases when the level of cognitive effort invested in a (conflict) task is high, even when subjects are not aware of systematic differences in difficulty between conditions (Diede & Bugg, 2017). Related, it has been shown recently that when agents are not aware of specific transitional rules in an implicit learning task, both the pupil and central ERP potentials (reminiscent of the mismatch negativity) may still signal surprise when statistical regularities in stimulus transitions are violated (Alamia et al., 2019; see also Meijs et al., 2018). Although intriguing, both tasks can be considered "implicit" (learning/conflict) tasks, because stimuli were always presented fully consciously and subjects were just not aware of differences in the probabilities of occurrence of specific stimuli. Therefore these effects cannot be directly compared to situation in which stimulus visibility is reduced.

Although we show that two different measures of the subjects arousal state reflect subjective prediction error responses, they may do so in potentially different ways, because correlations between pupil size and P3 amplitude were absent, in line with previous studies (Hong et al., 2014; Kamp & Donchin, 2015; Mückschel et al., 2017; Murphy et al., 2011). Recent animal studies have revealed a tight coupling between pupil diameter and neural responses in the noradrenergic locus coeruleus (Breton-Provencher & Sur, 2019; Joshi et al., 2016; Reimer et al., 2016; Liu et al., 2017; Varazzani et al., 2015), which is supported by recent human fMRI studies (de Gee et al., 2017; Murphy, O'Connell, et al., 2014). However, some (of these) studies also found unique contributions to pupil size in other subcortical regions, such as the cholinergic basal forebrain, dopaminergic midbrain, and the superior and inferior colliculi (de Gee et al., 2017; Joshi et al., 2016; Mridha et al., 2019; Reimer et al., 2016). Several lines of evidence reinforce a putative link between pupil diameter and the dopamine system, for example in patients with Parkinson's disease (Kringelbach et al., 2007; Manohar & Husain, 2015; Mathôt, 2018; Varazzani et al., 2015; Weinshenker & Schroeder, 2007). Similarly, the P3 has also been used as an electrophysiological correlate of feedback-evoked phasic catecholamine release in the cortex (Nieuwenhuis et al., 2005; Polich, 2007; Rangel-Gomez et al., 2013). Therefore, although these physiological markers of phasic arousal tend co-occur, they may reflect (partly) different processes (Eckstein et al., 2017; Kamp & Donchin, 2015). Understanding the relationship between changes in pupil dilation and the amplitude of P3 responses is important avenue for future research. Further, although the main goal of this study was testing the association between putative measures of central arousal state (pupil size/P3 responses) and subjective prediction error computation, exploratory analyses did not reveal any systematic relationship between other ERP components associated with feedback

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

processing, such as the feedback related negativity (FRN) (Cohen et al., 2011) and our variables of interest. Previously, using a probabilistic reversal learning task, we observed that the amplitude of the FRN was strongly linked to the signed prediction error variable ("objective prediction error") derived from reinforcement learning modeling (Correa et al., 2018). This relationship was strongly attenuated when feedback awareness was reduced. Future work is needed to explore in more detail the relationship between different ERP components (e.g. FRN, P3) and specific aspects of prediction error computation and how these may be differentially affected by levels of (feedback) awareness.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Alamia, A., VanRullen, R., Pasqualotto, E., Mouraux, A., & Zenon, A. (2019). Pupil-linked arousal responds to unconscious surprisal. *The Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.3010-18.2019

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*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bianco, R., Ptasczynski, L. E., & Omigie, D. (2020). Pupil responses to pitch deviants reflect predictability of melodic sequences. *Brain and Cognition*. https://doi.org/10.1016/j.bandc.2019.103621

Boldt, a., & Yeung, N. (2015). Shared Neural Markers of Decision Confidence and Error Detection. *Journal of Neuroscience*, *35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Bouret, S., & Sara, S. J. (2005). Network reset: A simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2005.09.002

Breton-Provencher, V., & Sur, M. (2019). Active control of arousal by a locus coeruleus GABAergic circuit. *Nature Neuroscience*. https://doi.org/10.1038/s41593-018-0305-z

Brown, S. B. R. E., van der Wee, N. J. A., van Noorden, M. S., Giltay, E. J., & Nieuwenhuis, S. (2015). Noradrenergic and cholinergic modulation of late ERP responses to deviant stimuli. *Psychophysiology*, *52*(12), 1620–1631. https://doi.org/10.1111/psyp.12544

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage*, *73*, 80–94. https://doi.org/10.1016/j.neuroimage.2013.01.054

Cohen, M. X., van Gaal, S., Ridderinkhof, K. R., & Lamme, V. a F. (2009). Unconscious errors enhance prefrontal-occipital oscillatory synchrony. *Frontiers in Human Neuroscience*, *3*(November), 54. https://doi.org/10.3389/neuro.09.054.2009

Cohen, M. X., Wilmes, K., & van de Vijver, I. (2011). Cortical electrophysiological network dynamics of feedback learning. *Trends in Cognitive Sciences*, *15*, 558–566. https://doi.org/10.1016/j.tics.2011.10.004

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), 1–13. https://doi.org/10.1038/s41598-018-31985-3

Correa, C. M. C., Noorman, S., Jiang, J., Palminteri, S., Cohen, M. X., Lebreton, M., & van Gaal, S. (2018). How the level of reward awareness changes the computational and electrophysiological signatures of reinforcement learning. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0457-18.2018

Critchley, H. D. (2005). Neural mechanisms of autonomic, affective, and cognitive integration. *Journal of Comparative Neurology*, *493*(March), 154–166. https://doi.org/10.1002/cne.20749

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*. 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*. 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 of the United States of America*, *111*, E618-25. https://doi.org/10.1073/pnas.1317557111

Dehaene, S., & Changeux, J. P. (2011). Experimental and Theoretical Approaches to

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Conscious Processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*, 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2

Diede, N. T., & Bugg, J. M. (2017). Cognitive effort is modulated outside of the explicit awareness of conflict frequency: Evidence from pupillometry. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000349

Doya, K. (2008). Modulators of decision making. *Nature Neuroscience*, *11*(March), 410–416. https://doi.org/10.1038/nn2077

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? In *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2016.11.001

Filipowicz, A. L., Glaze, C. M., Kable, J. W., & Gold, J. I. (2020). Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *Elife*, *9*, e57872.

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., & Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nature neuroscience*, *22*(12), 2066-2077.

Fleming, S., & Dolan, R. (2012). *The neural basis of metacognitive ability*. 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00443

Friedman, D., Hakerem, G., Sutton, S., & Fleiss, J. L. (1973). Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. *Electroencephalography and Clinical Neurophysiology*. https://doi.org/10.1016/0013-4694(73)90065-5

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. In *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196546

Glimcher, P. W. (2005). Indeterminacy in brain and behavior. *Annual Review of Psychology*, *56*, 25–56. https://doi.org/10.1146/annurev.psych.55.090902.141429

Gold, J., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: a quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*. https://doi.org/10.3758/BF03204445

Hommel, B. (2007). Consciousness and Control. *Journal of Consciousness Studies*, *14*(1), 155–176. https://doi.org/10.1006/ccog.2001.0521

Hong, L., Walz, J. M., & Sajda, P. (2014). Your eyes give you away: Prestimulus changes in pupil diameter correlate with poststimulus task-related EEG dynamics. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0091321

Jepma, M., Murphy, P. R., Nassar, M. R., Rangel-Gomez, M., Meeter, M., & Nieuwenhuis, S. (2016). Catecholaminergic Regulation of Learning Rate in a Dynamic Environment. *PLoS Computational Biology*, *12*(10). https://doi.org/10.1371/journal.pcbi.1005171

Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition. In *Trends in Cognitive Sciences*. 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

Kamp, S. M., & Donchin, E. (2015a). ERP and pupil responses to deviance in an oddball paradigm. *Psychophysiology*. https://doi.org/10.1111/psyp.12378

Kamp, S. M., & Donchin, E. (2015b). ERP and pupil responses to deviance in an oddball

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

paradigm. *Psychophysiology*. https://doi.org/10.1111/psyp.12378

Kiefer, M., Ansorge, U., Haynes, J. D., Hamker, F., Mattler, U., Verleger, R., & Niedeggen, M. (2011). Neuro-cognitive mechanisms of conscious and unconscious visual perception: From a plethora of phenomena to general principles. *Advances in Cognitive Psychology*, *7*, 55–67. https://doi.org/10.2478/v10053-008-0090-4

Kiesel, A., Kunde, W., & Hoffmann, J. (2007). Mechanisms of subliminal response priming. *Advances in Cognitive Psychology*, *3*(1), 307–315. https://doi.org/10.2478/v10053-008-0032-1

Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.12859

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. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0155574

Kringelbach, M. L., Jenkinson, N., Owen, S. L. F., & Aziz, T. Z. (2007). Translational principles of deep brain stimulation. *Nature Reviews. Neuroscience*, *8*(July), 623–635. https://doi.org/10.1038/nrn2196

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*. https://doi.org/10.1038/s41562-017-0107

Kunde, W., Reuss, H., & Kiesel, A. (2012). Consciousness and cognitive control. *Advances in Cognitive Psychology*, *8*(1), 9–18. https://doi.org/10.2478/v10053-008-0097-x

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*. https://doi.org/10.1016/j.cub.2017.02.026

Lamme, V. a. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*(3), 204–220. https://doi.org/10.1080/17588921003731586

Larsen, R. S., & Waters, J. (2018). Neuromodulatory correlates of pupil dilation. In *Frontiers in Neural Circuits*. https://doi.org/10.3389/fncir.2018.00021

Lebreton, M., Langdon, S., Slieker, M. J., Nooitgedacht, J. S., Goudriaan, A. E., Denys, D., van Holst, R. J., & Luigjes, J. (2018). Two sides of the same coin: Monetary incentives concurrently improve and bias confidence judgments. *Science Advances*. https://doi.org/10.1126/sciadv.aaq0668

Lempert, K. M., Chen, Y. L., & Fleming, S. M. (2015). Relating pupil dilation and metacognitive confidence during auditory decision-making. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0126588

Liao, H.-I., Yoneya, M., Kashino, M., & Furukawa, S. (2018). Pupillary dilation response reflects surprising moments in music. *Journal of Eye Movement Research*. https://doi.org/DOI: 10.16910/jemr.11.2.13

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 Reports*. https://doi.org/10.1016/j.celrep.2017.08.094

Maier, M. E., Ernst, B., & Steinhauser, M. (2019). Error-related pupil dilation is sensitive to the evaluation of different error types. *Biological Psychology*. https://doi.org/10.1016/j.biopsycho.2018.12.013

Manohar, S. G., & Husain, M. (2015). Reduced pupillary reward sensitivity in Parkinson's disease. *Npj Parkinson's Disease*, *1*(1), 20–23. https://doi.org/10.1038/npjparkd.2015.26

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., Tolias, A.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Meijs, E. L., Slagter, H. A., de Lange, F. P., & van Gaal, S. (2018). Dynamic interactions between top–down expectations and conscious awareness. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1952-17.2017

Meyniel, F., Sergent, C., Rigoux, L., Daunizeau, J., & Pessiglione, M. (2013). Neurocomputational account of how the human brain decides when to have a break. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 2641–2646. https://doi.org/10.1073/pnas.1211925110

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Perspective Confidence as Bayesian Probability : From Neural Origins to Behavior. *Neuron*, *88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039

Mridha, Z., Gee, J. W. de, Shi, Y., Alkashgari, R., Williams, J., Suminski, A., Ward, M. P., Zhang, W., & McGinley, M. J. (2019). Graded recruitment of pupil-linked neuromodulation by parametric stimulation of the vagus nerve. *BioRxiv*. https://doi.org/10.1101/2019.12.28.890111

Mückschel, M., Chmielewski, W., Ziemssen, T., & Beste, C. (2017). The norepinephrine system shows information-content specific properties during cognitive control – Evidence from EEG and pupillary responses. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2017.01.036

Murphy, P. R., Boonstra, E., & Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nature Communications*. https://doi.org/10.1038/ncomms13526

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*. https://doi.org/10.1111/j.1469-8986.2011.01226.x

Murphy, P. R., Vandekerckhove, J., & Nieuwenhuis, S. (2014). Pupil-Linked Arousal Determines Variability in Perceptual Decision Making. *PLoS Computational Biology*, *10*(9), e1003854. https://doi.org/10.1371/journal.pcbi.1003854

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

Nieuwenhuis, S, Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, a. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology*, *38*, 752–760. https://doi.org/10.1111/1469-8986.3850752

Nieuwenhuis, Sander, Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

O'Reilly, J. X., Schüffelgen, 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 of the United States of America*, 1–10. https://doi.org/10.1073/pnas.1305373110

Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of Psychophysiology*, *19*(4), 319–329. https://doi.org/10.1027/0269-8803.19.4.319

Parikh, V., Kozak, R., Martinez, V., & Sarter, M. (2007). Prefrontal Acetylcholine Release

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Controls Cue Detection on Multiple Timescales. *Neuron*. https://doi.org/10.1016/j.neuron.2007.08.025

Pineda, J. a, Foote, S. L., & Neville, H. J. (1989). Effects of locus coeruleus lesions on auditory, long-latency, event-related potentials in monkey. *The Journal of Neuroscience*, *9*(1), 81–93.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. In *Clinical Neurophysiology* (Vol. 118, Issue 10, pp. 2128–2148). https://doi.org/10.1016/j.clinph.2007.04.019

Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2011.00115

Raisig, S., Welke, T., Hagendorf, H., & van der Meer, E. (2010). I spy with my little eye: Detection of temporal violations in event sequences and the pupillary response. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2010.01.006

Rangel-Gomez, M., Hickey, C., van Amelsvoort, T., Bet, P., & Meeter, M. (2013). The Detection of Novelty Relies on Dopaminergic Signaling: Evidence from Apomorphine's Impact on the Novelty N2. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0066469

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*. https://doi.org/10.1038/ncomms13289

Richer, F., & Beatty, J. (1987). Contrasting Effects of Response Uncertainty on the Task-Evoked Pupillary Response and Reaction Time. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.1987.tb00291.x

Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a Statistical Computation in the Human Sense of Confidence. *Neuron*. https://doi.org/10.1016/j.neuron.2016.03.025

Shalgi, S. (2012). Is any awareness necessary for an Ne? *Frontiers in Human Neuroscience*, *6*(May), 1–15. https://doi.org/10.3389/fnhum.2012.00124

Steinhauer, S., & Zubin, J. (1982). VULNERABILITY TO SCHIZOPHRENIA: INFORMATION PROCESSING IN THE PUPIL AND EVENT-RELATED POTENTIAL. In *Biological Markers in Psychiatry and Neurology*. https://doi.org/10.1016/b978-0-08-027987-9.50042-1

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. *Nature Reviews. Neuroscience*, *6*(May), 363–375. https://doi.org/10.1038/nrn1666

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in Cognitive Sciences*, 1–9. https://doi.org/10.1016/j.tics.2014.02.009

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*, 1–11. https://doi.org/10.1038/ncomms14637

van Gaal, S., & Lamme, V. a. F. (2012). Unconscious High-Level Information Processing: Implication for Neurobiological Theories of Consciousness. *The Neuroscientist*, *18*, 287–301. https://doi.org/10.1177/1073858411404079

van Gaal, Simon, de Lange, F. P., & Cohen, M. X. (2012). The role of consciousness in cognitive control and decision making. *Frontiers in Human Neuroscience*, *6*(May), 1–15. https://doi.org/10.3389/fnhum.2012.00121

Van Slooten, J. C., Jahfari, S., Knapen, T., & Theeuwes, J. (2018). How pupil responses track value-based decision-making during and after reinforcement learning. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1006632

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.25.164962; this version posted June 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dopamine Neurons in the Reward/Effort Trade-Off: A Direct Electrophysiological Comparison in Behaving Monkeys. *Journal of Neuroscience*, *35*(20), 7866–7877. https://doi.org/10.1523/JNEUROSCI.0454-15.2015

Vincent, P., Parr, T., Benrimoh, D., & Friston, K. J. (2019). With an eye on uncertainty: Modelling pupillary responses to environmental volatility. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1007126

Weinshenker, D., & Schroeder, J. P. (2007). There and back again: A tale of norepinephrine and drug addiction. *Neuropsychopharmacology*, *32*(7), 1433–1451. https://doi.org/10.1038/sj.npp.1301263

Wetzel, N., Buttelmann, D., Schieler, A., & Widmann, A. (2016). Infant and adult pupil dilation in response to unexpected sounds. *Developmental Psychobiology*. https://doi.org/10.1002/dev.21377

Wyart, V., & Koechlin, E. (2016). Choice variability and suboptimality in uncertain environments. In *Current Opinion in Behavioral Sciences* (Vol. 11, pp. 109–115). https://doi.org/10.1016/j.cobeha.2016.07.003

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. In *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2011.0416

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*, 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zhao, S., Chait, M., Dick, F., Dayan, P., Furukawa, S., & Liao, H. I. (2019). Pupil-linked phasic arousal evoked by violation but not emergence of regularity within rapid sound sequences. *Nature Communications*. https://doi.org/10.1038/s41467-019-12048-1

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness and Cognition*, *27*, 246–253. https://doi.org/10.1016/j.concog.2014.05.012