Pupil-linked arousal modulates precision of stimulus representation in cortex

Abbreviated title: Arousal modulates sensory precision

Laura S. Geurts[1], Sam Ling[2,3], & Janneke F.M. Jehee[1]

[1]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, 6525 EN Nijmegen, the Netherlands;

[2]Department of Psychological and Brain Sciences, Boston University, Boston, Massachusetts 02215; [3]Center for Systems Neuroscience, Boston University, Boston, Massachusetts 02215

Corresponding author: Janneke Jehee, janneke.jehee@donders.ru.nl

Number of pages: 37

Number of figures: 6

Number of tables: 0

Number of words:

- Abstract: 251
- Introduction: 659
- Discussion: 1067

Conflict of interest statement

The authors declare no competing financial interests.

Acknowledgments

We thank P. Gaalman for MRI support. This work was funded by ERC Starting Grant No. 677601 to J.F.M. Jehee, and S. Ling was supported by NIH Grant EY028163.

ABSTRACT

Neural responses are naturally variable from one moment to the next, even when the stimulus is held constant. What factors might underlie this variability in neural population activity? We hypothesized that spontaneous fluctuations in the cortical stimulus representation are created by changes in arousal state. We tested the hypothesis using a combination of fMRI, probabilistic decoding methods and pupillometry. Human participants (20 female, 12 male) were presented with gratings of random orientation. Shortly after viewing the grating, participants reported its orientation and gave their level of confidence in this judgment. Using a probabilistic fMRI decoding technique, we quantified the precision of the stimulus representation in the visual cortex on a trial-by-trial basis. Pupil size was recorded and analyzed to index the observer's arousal state. We found that the precision of the cortical stimulus representation, reported confidence, and variability in the behavioral orientation judgments varied from trial to trial. Interestingly, these trial-by-trial changes in cortical and behavioral precision and confidence were linked to pupil size and its temporal rate of change. Specifically, when the cortical stimulus representation was more precise, the pupil dilated more strongly prior to stimulus onset and remained larger during stimulus presentation. Similarly, stronger pupil dilation during stimulus presentation was associated with higher levels of subjective confidence, a secondary measure of sensory precision, as well as improved behavioral performance. Taken together, our findings support the hypothesis that spontaneous fluctuations in arousal state modulate the fidelity of the stimulus representation in the human visual cortex, with clear consequences for behavior.

SIGNIFICANCE STATEMENT

The fidelity of our sensory experiences varies from moment to moment. For example, we sometimes fail to recognize a friend in a crowd or mistake them for someone else. What determines the quality of human sensation and perception? In this study, we investigated whether fluctuations in alertness might play a role. We recorded brain activity while participants viewed images and reported both what they had seen and how confident they felt in this judgment. We discovered that a spontaneous change in alertness impacts the fidelity of information processing in the visual brain as well as reported levels of confidence and behavioral performance. These findings provide new insight into the mechanisms that underlie spontaneous changes in sensory information processing in the human brain.

INTRODUCTION

Neural and behavioral responses are rarely constant over time – not even for repeated presentations of the same stimulus (see e.g. Faisal et al., 2008 for a review). While numerous processes underlie these apparent fluctuations in brain and behavior, spontaneous changes in arousal state likely play an outsized role in this variability. Arousal refers to a state of physiological alertness or readiness, mediated by brainstem neuromodulatory systems, with wide-spread influences on neural and physiological activity. Previous work has shown that arousal modulates overall activity in the visual cortex (Livingstone and Hubel, 1981; Reimer et al., 2014; Vinck et al., 2015; Roth et al., 2020) and even the retina (Schröder et al., 2020), but whether it influences the quality of sensory information processing remains unknown. Here, we propose that spontaneous fluctuations in arousal modulate the fidelity of stimulus representations in cortex. More specifically, we hypothesize that arousal enhances the precision with which sensory information is represented in neural population activity in the visual cortex.

To test this hypothesis, we measured cortical activity with functional magnetic resonance imaging (fMRI) while participants performed a perceptual judgment task, and recorded pupil size as an index of arousal. Pupil size is commonly used as an indicator of arousal state, motivated by the tight links between pupil dilator muscles and the locus coeruleus-norepinephrine (LC-NE) system (see Mathôt, 2018, for a review), which is believed to play a central role in arousal (Moruzzi and Magoun, 1949; Aston-Jones and Cohen, 2005; Sara, 2009). Indeed, the relationship between pupil size and arousal has been demonstrated in both neurophysiological (Aston-Jones and Cohen, 2005; Varazzani et al., 2015; Joshi et al., 2016) and neuroimaging studies (Murphy et al., 2014). Activity in the LC-NE system is not only linked to pupil size per se, but is particularly tightly coupled with rapid (phasic) changes in pupil size (Reimer et al., 2016). In recent years, researchers have therefore started using the rate of change in pupil size, quantified as the first derivative or slope of the pupil signal, as an (additional) measure of

arousal (e.g. de Gee et al., 2020; Podvalny et al., 2021; Pfeffer et al., 2022). In this study, we consider both pupil size and its rate of change (slope).

To quantify the quality of the stimulus representation in the visual cortex, we applied a probabilistic decoding technique (van Bergen et al., 2015; van Bergen and Jehee, 2018). This method decodes stimulus information from a pattern of cortical activity as a probability distribution over all possible stimuli – on a per-trial basis. Importantly, the width of the decoded distribution provides a metric of the uncertainty associated with the cortical stimulus representation: the wider the decoded distribution, the wider the range of stimuli that are consistent with the activity pattern. Vice versa, a narrow decoded distribution suggests that only a few stimuli are likely to have triggered the given response pattern; in other words, the cortical representation of the stimulus is very precise. Previous work has shown that decoded uncertainty provides a reliable measure of the quality of the cortical representation of sensory information (van Bergen et al., 2015; van Bergen and Jehee, 2019; Li et al., 2021; Geurts et al., 2022; Chetverikov and Jehee, 2023). Because sensory uncertainty has been linked to the participant's self-reported levels of confidence about their perceptual decisions (Geurts et al., 2022), we considered reported confidence as a secondary measure of sensory uncertainty. To assess arousal's impact on behavior, we also quantified the precision of the participant's behavioral responses across trials.

Using these methods, we found evidence to suggest that spontaneous fluctuations in pupil-linked arousal state modulate the fidelity of stimulus representations in the human visual cortex. Specifically, we observed that decoded uncertainty, reported levels of confidence and behavioral precision vary from moment to moment and are linked to both pupil size and its rate of change. These results support the hypothesis that arousal plays a role in modulating the quality of sensory representations in the human visual cortex.

MATERIALS AND METHODS

Participants

Thirty-two healthy adult volunteers (20 female, 12 male, age: 19-31 years) with normal or corrected-to-normal vision participated in this study, which was approved by the local medical ethics review committee (CMO Arnhem-Nijmegen, the Netherlands). All participants provided informed written consent and received monetary compensation for their participation. The sample size (N = 32) was based on a power calculation for detecting a reliable correlation between decoded uncertainty and behavioral variability using data from a previous study with a similar design (van Bergen et al., 2015; power = 0.8; $\alpha = 0.05$). Participants were included based on their ability to perform the task, which was assessed in a separate behavioral training session prior to the experimental sessions.

Imaging data acquisition

The MRI data were collected using a Siemens 3T MAGNETOM PrismaFit scanner and a 32-channel head coil at the Donders Centre for Cognitive Neuroimaging in Nijmegen, the Netherlands. The data were analyzed previously for a different purpose (Geurts et al., 2022). Each scan session started with the collection of T1-weighted image (3D MPRAGE; repetition time (TR): 2300 ms; inversion time (TI): 1100 ms; echo time (TE): 3 ms; flip angle: 8 degrees; field of view (FOV), 256 mm × 256 mm; 192 sagittal slices; 1-mm isotropic voxels) and B0 field inhomogeneity maps (TR: 653 ms; TE: 4.92 ms; flip angle: 60 degrees; FOV: 256 mm × 256 mm; 68 transversal slices; 2-mm isotropic voxels; interleaved slice acquisition). Functional MRI data were acquired using a multi-band accelerated gradient-echo EPI sequence, with 68 transversal slices covering the whole brain (TR: 1500 ms; TE: 38.60 ms; flip angle: 75 degrees; FOV: 210 mm × 210 mm; 2-mm isotropic voxels; multiband acceleration factor: 4; interleaved slice acquisition).

Pupil data acquisition

Pupillometry data were acquired using an SR Research Eyelink 1000 system. Pupil size was sampled at 1 kHz. Pupil recordings were collected for 62 out of 64 sessions, and only partially (4-12 runs out of a total of 10-13) for 11 of these sessions, due to technical difficulties.

Experimental design

Participants performed an orientation estimation task (Figure 1) inside the MRI scanner. They were instructed to maintain fixation on a black-and-white bullseye target (radius: 0.375 degrees) presented at the center of the screen throughout each run. Runs consisted of 20 trials each (trial duration: 16.5 s, inter-trial interval: 1.5 s) and started and ended with a fixation period (duration: 4.5 and 16 s, respectively). Each trial began with the presentation of an orientation stimulus (duration: 1.5 s), followed by a 6-s retention interval, and two 4.5-s response windows in which observers were prompted to report the orientation of the viewed grating and indicate their level of confidence in this orientation response. The stimuli were counterphasing sinusoidal gratings (contrast: 10%; spatial frequency: one cycle per degree; randomized spatial phase; 2-Hz sinusoidal contrast modulation), presented inside an annulus around fixation (inner radius: 1.5 degrees; outer radius: 7.5 degrees; contrast linearly decreasing over the inner and outer 0.5 degrees of the annulus). Stimulus orientations were drawn pseudorandomly from a uniform distribution (0-179 degrees) to ensure an approximately even sampling of orientation within any given run. During the first response window, participants reported the orientation of the viewed grating by rotating a black bar (length: 2.8 degrees; width: 0.1 degrees; contrast: 40%; initial orientation randomized across trials) presented at the center of the screen. During the second response window, participants indicated their confidence in this orientation judgment by sliding a white dot over a circular scale. The scale was a black bar of increasing width (contrast: 40%; bar width: 0.1–0.5 degrees, linearly increasing) that was wrapped around fixation (radius: 1.4 degrees). The mapping between confidence level and scale width (i.e., whether the narrow end of the scale indicated high or low confidence) was counterbalanced across participants, and the

orientation and direction of the scale (i.e., whether the width increased in clockwise or counterclockwise direction), as well as the dot's starting position, were randomized across trials. For both response windows, the response bar (or scale) faded linearly over the last 1 s of the response window to indicate the approaching end of this window, and participants responded using two buttons (one for clockwise and one for counterclockwise rotation) on an MRI-compatible button box. Each trial was preceded by the fixation bullseye briefly turning black (duration: 0.1 s, timing: -0.5 s relative to stimulus onset) as a cue to stimulus onset. Participants performed a total of 22-26 task runs inside the scanner, divided over two sessions on separate days, and extensively practiced the task in separate behavioral sessions prior to the experiment (2-4 hours in total).

Each scan session also included one or two functional localizer runs, in which flickering checkerboard stimuli (contrast: 100%, flicker frequency: 10 Hz, check size: 0.5 degrees) were presented within the same annulus as the orientation stimuli. Checkerboard stimuli were presented in seven 12-s blocks interleaved with fixation blocks of equal duration. In a separate scan session, retinotopic maps of the visual cortex were acquired using standard retinotopic mapping procedures (Sereno et al., 1995; Deyoe et al., 1996; Engel et al., 1997).

Visual stimuli were generated by a Macbook Pro computer using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007), and displayed via a luminance-calibrated projector (EIKI LC-XL100; screen resolution: 1,024 × 768 pixels; refresh rate: 60 Hz) on a rear-projection screen, which the participants viewed via a mirror mounted on the head coil.

Preprocessing of MRI data

Preprocessing procedures for functional imaging data are also described in Geurts et al. (2022), and reproduced here for convenience. Motion correction was performed with respect to the middle volume of the middle run of each session (the motion correction template) with FSL's MCFLIRT (Jenkinson et al., 2002). The motion-correction template was corrected for distortions due to B0 field

inhomogeneities using the acquired field maps, and registered to a high-resolution anatomical (T1-weighted) image acquired in the same session using epi_reg within FSL's FLIRT (Jenkinson and Smith, 2001). For co-registration of data across sessions, we created participant-specific anatomical templates by combining the anatomical reference images from the two separate sessions using Freesurfer's mri_robust_template (Reuter et al., 2012), to which the single-session anatomical images were registered. All transformations were then combined and applied to the raw data. To remove slow drifts in the MRI signal, the transformed data were temporally filtered using FSL's nonlinear high-pass filter (Jenkinson et al., 2012) with a sigma of 24 TRs (two trials), which corresponds to a cut-off of around 83 s. Residual motion effects were removed from the data through linear regression, using a set of 24 motion regressors derived from the motion parameters estimated by MCFLIRT.

The region of interest (ROI) for decoding (bilateral V1, V2, and V3, combined) was identified on the reconstructed cortical surface, obtained with Freesurfer's cortical reconstruction algorithm (Dale et al., 1999), using single-participant retinotopic maps (see *Experimental design*). For further analysis, we selected the 2000 voxels within this ROI that were most strongly activated by the functional localizer stimulus while surviving a lenient statistical threshold of $p<0.01$, uncorrected. Voxel selection was performed for each participant individually in native space. Each voxel's timeseries were z-normalized with respect to corresponding trial time points in the same run. Finally, we obtained single-trial activation patterns by adding a 4.5-s temporal shift (to account for the hemodynamic delay) and then averaging over the first 3 s of each trial. Importantly, this time window excludes activity from the behavioral response window.

Decoding algorithm

To quantify the trial-by-trial precision of stimulus representations in visual cortex, we applied a generative-model based probabilistic decoding algorithm (van Bergen et al., 2015; van Bergen and Jehee, 2018) to our data (see *Preprocessing of MRI data* for voxel selection criteria). We provide a concise description of the decoding algorithm here, and refer the interested reader to previous

publications for further detail (van Bergen et al., 2015; van Bergen and Jehee, 2018). The decoding model describes the across-trial distribution of activation patterns as a multivariate normal, centered around a stimulus-specific mean that describes the tuning function of each voxel. Tuning functions were modeled as a linear combination $\mathbf{W}\mathbf{f}(s)$ of eight bell-shaped basis functions $\mathbf{f}(s) = [f_1(s), ..., f_8(s)]^{\mathrm{T}}$, each centered on a different orientation (Brouwer and Heeger, 2009). The basis functions are described as follows:

$$\mathrm{f_k}\,(\mathrm{s}) \;=\; \max\left(0,\, \cos\left(2\pi\frac{\mathrm{s}-\phi_\mathrm{k}}{180}\right)\right)^5$$

in which s is the orientation of the presented stimulus and ϕ_k is the center of the kth basis function. Basis functions were spaced equally across the full orientation space (0-179 degrees) with the first one centered at zero. The basis functions are weighted by coefficients $\mathbf{W}$, with W_{ik} the contribution of the kth basis function to the tuning function of the ith voxel.

The variance around the stimulus-dependent mean activation pattern (i.e., the multivariate tuning function) is modeled by the covariance matrix:

$$\mathbf{\Omega} = \rho\boldsymbol{\tau}\boldsymbol{\tau}^{\mathrm{T}} + (1-\rho)\mathrm{diag}\left(\boldsymbol{\tau}^{\mathbf{2}}\right) + \sigma^2\mathbf{W}\mathbf{W}^{\mathrm{T}}$$

The first component of the covariance matrix $\rho\boldsymbol{\tau}\boldsymbol{\tau}^{\mathrm{T}}$ models global fluctuations shared between all voxels in the ROI. The second component$(1-\rho)\mathrm{diag}\left(\boldsymbol{\tau}^{\mathbf{2}}\right)$, with $\boldsymbol{\tau}^2 = \left[\tau_i^2\right]^{\mathrm{T}}$, describes independent, voxel-specific variability (with variance τ_i^2 for voxel i). The relative contributions of these two components are given by ρ. The third component $\sigma^2\mathbf{W}\mathbf{W}^{\mathrm{T}}$ reflects variance (with magnitude σ^2) shared between voxels with similar orientation preference (given by $\mathbf{W}\mathbf{W}^{\mathrm{T}}$).

The voxel tuning functions and covariance matrix together model the generative distribution of activation patterns:

$$p(\mathbf{b}|s,\theta) = \mathrm{N}(\mathbf{W}\mathbf{f}(s), \mathbf{\Omega})$$

which is a multivariate normal distribution with mean $\mathbf{W}\mathbf{f}(s)$ and covariance $\mathbf{\Omega}$. $\theta = \{\mathbf{W}, \rho, \boldsymbol{\tau}, \sigma\}$ are the generative model's parameters.

For model training and testing, a leave-one-run-out cross-validation procedure was used to prevent double-dipping. The model's parameters were estimated on a dataset consisting of data from all but one task run. The trained model was then tested on the held-out run, and this procedure was repeated until all runs had served as a test run exactly once. The parameters were estimated in two steps: the coefficients $\mathbf{W}$ were first estimated by ordinary least squares regression, and then the covariance parameters ($\rho, \boldsymbol{\tau}, \sigma$) were estimated through numerical likelihood maximization (see van Bergen et al., 2015 for further details regarding model estimation procedures).

Model testing ('decoding') consisted of calculating a posterior distribution $p(s|\mathbf{b}, \hat{\theta})$ over stimulus orientation s, conditioned on the response pattern $\mathbf{b}$ and estimated parameters $\hat{\theta}$. The posterior distribution is given by Bayes' rule:

$$p(s|\mathbf{b}, \hat{\theta}) = \frac{p(\mathbf{b}|s, \hat{\theta})p(s)}{\int p(\mathbf{b}|s, \hat{\theta})p(s)}$$

The stimulus prior $p(s)$ was flat, reflecting the uniform stimulus distribution used in the experiment. The normalization constant $\int p(\mathbf{b}|s, \hat{\theta})p(s)$ was computed numerically. The circular mean of the decoded distribution was taken as the decoder's estimate of the presented orientation ('decoded orientation') and the squared circular standard deviation quantified the uncertainty in this estimate ('decoded uncertainty').

Preprocessing of pupil data

Blinks and saccades were identified using the Eyelink software. Data recorded during saccades or less than 250 ms before (after) blink onset (offset) were removed. Missing or removed data were linearly interpolated. Data interpolated over more than 1000 ms were removed at the end of the preprocessing

procedure. If more than 50 % of the data in a given trial were missing, the trial was excluded from pupil data analyses.

The pupil's responses to blinks and saccades were estimated and regressed out using a deconvolution approach developed by Knapen et al. (2016), and implemented by Urai et al. (2017). Specifically, the shape of blink- and saccade-triggered pupil responses was first estimated by fitting a finite impulse response (FIR) model to the data of each participant. The estimated response was then used to create a regressor in a linear regression analysis, and blink- and saccade-related effects were removed (separately for each run). Data were subsequently low-pass filtered using a third-order Butterworth filter with a cut-off of 4 Hz, and downsampled to 100 Hz. Global effects in the data (cf. Knapen et al., 2016) were removed by fitting an exponential function to each run and using the residuals of this fit in subsequent analyses. Finally, pupil size was z-normalized per session to correct for differences in camera and lighting position between sessions.

We considered both the pupil size time series and its first derivative as indices of arousal. To estimate the derivative, we used a moving window with a width of 500 or 1000 ms. Within this window, we fitted a linear function to the pupil time series. We took the slope of this fitted function as a measure of the rate of change in pupil size at the time point on which the moving window was centered.

Preprocessing of behavioral data

The error in the observer's orientation response on a given trial was calculated as the acute-angle difference between the presented and reported orientation. Participants generally performed well on the task, with a mean absolute orientation response error of 5.81 ± 1.29 degrees (mean ± s.d. across observers). To correct for orientation-dependent biases in the response (shifts in mean response), two fourth-degree polynomials modeling response error as a function of stimulus orientation were fit to each participant's data (van Bergen et al., 2015; Geurts et al., 2022). The first polynomial was fit to trials with presented stimulus orientation between 0 and 89 degrees and the second to trials with the

presented stimulus between 90 and 179 degrees. The residuals of these fits ('bias-corrected behavioral responses') were used in subsequent analyses. Trials on which the bias-corrected response was more than three standard deviations away from zero were marked as guesses and excluded from all further analyses (0.91 ± 0.31 percent of all trials; mean ± s.d. across observers). Confidence ratings were z-scored per session to correct for potential between-participant or between-session differences in usage of the confidence scale. We excluded trials on which observers did not finish adjusting their orientation and/or confidence response before the end of the response window (2.75 ± 2.40 percent of all trials; mean ± s.d. across observers).

Statistical procedures

To benchmark our decoding approach, we quantified orientation decoding performance by calculating the circular equivalent of Pearson's correlation coefficient between the presented and decoded orientation across trials and for each individual observer. To quantify the effect at the group level, the single-observer correlation coefficients were Fisher-transformed, and a weighted average was computed. The weight for the correlation coefficient of observer i was calculated as $w_i = \frac{1}{v_i}$, where v_i corresponds to the variance of the Fisher-transformed correlation coefficient (Hedges and Olkin, 1985). This variance is given by $v_i = \frac{1}{n_i - 3}$, in which n_i represents the number of trials. Statistical significance of the group-averaged correlation coefficients was assessed using a Z-test. The results from this analysis, as well as the correlation between decoded uncertainty and distance to the nearest cardinal axis and behavioral variability (see below), were also reported as benchmarks in a previous study (Geurts et al., 2022, Extended Data Figure 2).

To assess whether decoded uncertainty predicts behavioral variability, trials were divided into ten bins of increasing uncertainty for each individual observer. Mean decoded uncertainty was computed across all trials in each bin and behavioral variability was quantified as the squared circular standard deviation of the (bias-corrected) behavioral errors in the bin. Multiple linear regression was performed

to calculate the partial correlation coefficient for the relationship between decoded uncertainty and behavioral variability at the group level, with separate intercepts for each observer. Statistical significance was assessed by means of a t-test. We performed a number of control analyses, in which we varied several analysis parameters. First, because there is no principled way for determining the number of bins to use in this analysis, we ran two additional analyses using five or fifteen (instead of ten) uncertainty bins per participant. Second, because the strength of the link between decoded uncertainty and behavioral variability could vary across individuals, we performed a linear mixed-effects analysis (using MATLAB's *lmefit* function) in which both the intercept and the slope were modeled as random effects. This is in contrast to the multiple linear regression approach described above, which assumes that the intercepts are random variables while the slope is fixed. Third, to assess the influence of extreme values, we performed the analysis on different subsets of the data. The first subset excluded all trials for which the standard deviation of the decoded distribution was larger than 45 degrees. In the second subset, we simply excluded from our analyses the observer who gave rise to the data point in the top-right corner of Figure 2B. Statistical significance was assessed by means of t-tests on the partial correlation coefficients (r) or estimated slope (β) for the multiple regression and linear mixed effects analyses, respectively. In the linear mixed effects analysis, Satterthwaite approximation was used to estimate the effective degrees of freedom.

The relationship between pupil size and decoded uncertainty or reported confidence was tested in two different ways. For the first set of analyses, we divided trials into three bins per observer, based on the level of decoded uncertainty (or reported confidence), and compared pupil size between the highest and lowest bin. We computed t-values to quantify the difference between these bins. T-values were computed for each observer individually, and then averaged across observers. The analysis was performed for each time point within the window of interest; that is, from 1.5 s before stimulus onset until 1.5 s after stimulus offset. To assess statistical significance, threshold-free cluster enhancement (TFCE; Smith and Nichols, 2009) and permutation testing (1000 permutations) were used. The family-

wise error rate (FWER) was controlled by comparing the true single-timepoint TFCE scores against the null distribution of the maximum TFCE score across time obtained through data permutation (Nichols and Hayasaka, 2003).

In the second set of analyses, we quantified the relationship between pupil size (or slope) and decoded uncertainty (or reported confidence) on a trial-by-trial basis. To do so, we computed Spearman's correlation coefficient for each data point in the pupil time series, from 1.5 s before stimulus onset until 1.5 s after stimulus offset. Spearman's correlation coefficient was used because there is no a-priori reason to assume that the relationship should be linear. Correlation coefficients were computed for each observer individually and group-level correlation coefficients were calculated following similar procedures as for orientation decoding performance. Specifically, individual correlation coefficients were Fisher transformed and a weighted average was computed with weights $w_i = \frac{1}{v_i}$ (Hedges and Olkin, 1985). For Spearman's correlation coefficient, v_i is given by $v_i = \frac{1.06}{n_i - 3}$, with n_i representing the number of trials for observer i (Fieller and Pearson, 1961). As before, statistical significance was assessed using TFCE and permutation testing (1000 permutations), and the FWER was controlled by comparing against the null distribution of the maximum TFCE-value across time points.

For the visualizations in Figure 4D, pupil size was averaged over the stimulus presentation window on a trial-by-trial basis. Spearman correlation coefficients between decoded uncertainty and (mean) pupil size were computed per observer, Fisher-transformed and averaged as described above. Z-tests were used to assess significance both at the group level and for the example observer.

The relative effect size of arousal state on changes in uncertainty was determined as follows. Because the effect size cannot be determined directly from the observed relationship between arousal and uncertainty (decoded uncertainty reflects not only neural but also many non-neural sources of noise, including from the fMRI scanner), we instead relied on an indirect approach and compared the effect of pupil-linked arousal with that of a manipulation of stimulus orientation to acquire an understanding

of its relative contribution. The impact of stimulus orientation on decoded uncertainty was quantified by computing the Spearman correlation coefficient between decoded uncertainty and the distance between the presented stimulus orientation and the nearest cardinal axis (cf. Geurts et al. 2022, Extended Data Figure 2B). Correlation coefficients were computed per observer, Fisher transformed and then averaged across observers as described above. The impact of pupil-linked arousal was summarized by first averaging pupil size over the stimulus presentation window on a per trial basis (cf. Figure 4D), and then computing and averaging the Spearman correlation coefficients between these trial-by-trial values and decoded uncertainty as described in the previous paragraph. Effect sizes were defined as the absolute group-averaged Spearman correlation coefficient.

Code accessibility

All custom analysis code is available from the corresponding author upon request. Code for the probabilistic decoding algorithm can be found at https://github. com/jeheelab/.

RESULTS

Do spontaneous fluctuations in arousal modulate the cortical representation of the stimulus? To address this question, we presented thirty-two human observers with oriented gratings while simultaneously measuring pupil size and recording their brain activity with fMRI. Observers reported the orientation of the grating and rated their level of confidence in this orientation judgment (Figure 1). To quantify the degree of imprecision in the stimulus representation in visual cortex (areas V1, V2, and V3 combined), we used a probabilistic decoding technique (van Bergen et al., 2015; van Bergen and Jehee, 2018). This technique computes a probability distribution over stimulus orientation for each trial of cortical activity (Figure 2A, top panel). The width of the decoded distribution reflects the degree of uncertainty contained in the pattern of activity. We refer to this metric as 'decoded uncertainty'. To measure arousal, we relied on pupil recordings (Figure 2A, middle panel). That is, pupil size is an established indicator of arousal state (e.g. Aston-Jones and Cohen, 2005; Gilzenrat et al., 2010; Reimer et al., 2014; Vinck et al., 2015; de Gee et al., 2017, 2020; Urai et al., 2017; Pfeffer et al., 2022). Both physiological and neuroimaging studies have linked changes in pupil dilation to activity in the locus coeruleus (LC) and the release of noradrenaline (NE) (Aston-Jones and Cohen, 2005; Murphy et al., 2014; Varazzani et al., 2015; Joshi et al., 2016). Previous work suggests that while overall pupil size is influenced by multiple factors, rapid (phasic) changes in pupil size more specifically track activity in the LC-NE system (Reimer et al., 2016). To quantify arousal, we therefore considered both pupil size and the first derivative ('slope') of its timeseries, which is specifically sensitive to changes in pupil size. Note that there is an inherent link between (changes in) pupil slope and pupil size: a large (positive) pupil slope at any given moment in time should be linked to an increase in pupil size moments later. For this reason, a true change in arousal state should be reflected in the pupil signal via an effect on slope followed by one on size (albeit that the two measures need not be equally sensitive; Reimer et al., 2016).

Decoded distributions reflect presented stimulus and behavioral imprecision

To benchmark our decoding approach, we first tested how closely the decoder's orientation estimate (the mean of the decoded distribution) matched the presented orientation on a trial-by-trial basis. We computed the circular correlation coefficient between the decoded and presented orientations for each participant individually and then averaged the coefficients. This analysis revealed that the decoded and presented orientations were significantly correlated ($z = 83.58$, $p < 0.001$, $r = 0.60$, 95% CI = [0.58, 0.61], see also Extended Data Figure 2A in Geurts et al., 2022).

Having established that the presented orientation can reliably be extracted from cortical activity, we next asked whether the width of the decoded distribution is a meaningful measure of the degree of imprecision in the cortical stimulus representation. That is, a more precise representation in cortex should result in more precise (less variable) behavior. Is decoded width linked to behavioral variability, suggesting that it reflects the quality of the underlying neural representation? To address this question, we divided trials into bins of increasing distribution width (ten bins per participant), calculated the variance of the behavioral orientation estimates in each bin (Figure 2A, bottom panel), and quantified their relationship via a regression analysis. Replicating previous studies (van Bergen et al., 2015; Chetverikov and Jehee, 2023), this revealed a significant link between the width of the posterior distribution and behavioral variability (Figure 2B; $t(287) = 2.30$, $p = 0.011$, $r = 0.13$, 95% CI = [0.019, 0.25]). Specifically, the broader the distribution's width, the more variable the observers´ behavioral orientation estimates were. Control analyses, in which we varied the number of uncertainty bins, used two different statistical models (mixed versus fixed effects), and analyzed various subsets of the data (see Methods for details), showed that these results are fairly robust to changes in analysis parameters (see Figure 3 for data and statistics). Taken together, this suggests that posterior width provides a reliable measure of the degree of uncertainty contained in neural activity. Interestingly, it also shows that the imprecision in the cortical representation is not constant over trials, but rather varies from one trial to the next, with clear consequences for behavior.

Pupil-linked arousal reliably predicts decoded uncertainty

Our analyses revealed that uncertainty fluctuates considerably from one trial to the next. What processes might underlie these spontaneous changes in neural activity? Here, we hypothesize that fluctuations in sensory uncertainty might be linked to arousal state. That is, given that arousal is a physiological state of alertness with effects on neural activity (Livingstone and Hubel, 1981; Reimer et al., 2014; Vinck et al., 2015), we reasoned that higher levels of arousal might lead to better stimulus representations in cortex, and hence, lower stimulus uncertainty. To measure arousal, we recorded pupil size while participants performed the task in the scanner. We predicted that the size of the pupil should vary across trials and be linked to uncertainty. Specifically, sensory uncertainty should decrease when the pupil dilates, indicating higher levels of arousal. To index arousal state, we considered both pupil size and the slope of the pupil timeseries, which quantifies the rate of change in pupil size. Because we were interested in the effects of arousal on sensory uncertainty, we focused on pupil size and dilation just before, during and just after stimulus presentation (Figure 4A). Note that while a change in pupil size alone can affect retinal resolution (which might, in turn, affect downstream activity), the direction of this effect runs opposite to what we predict here, as retinal image quality is reduced for larger pupils due to spherical aberrations (Campbell and Gregory, 1960; Campbell and Green, 1965).

To test the link between pupil-linked arousal and representational fidelity, we first divided all trials of each individual participant into three equal-sized bins of increasing uncertainty, computed for each point in time the mean pupil size across all trials in each bin, and then compared between the lowest and highest uncertainty bin. Paired t-tests revealed a significant difference in pupil size between high and low uncertainty trials starting just before cue onset and lasting until at least 1.5 s after stimulus offset (Figure 4B, left panel; permutation tests, all $p<0.05$, FWER-corrected). Thus, the size of the pupil was larger when uncertainty in cortex was low and the stimulus representation was more precise.

Interestingly, the effect started well before stimulus onset, suggesting that altered arousal state led to the change in the cortical representation of the stimulus.

We next asked if decoded uncertainty is also linked to pupil size on a per trial basis. To address this question, we computed the trial-by-trial correlation coefficient between decoded uncertainty and pupil size (calculated separately for each time point). We did this first for each individual observer and then averaged across observers (see Methods for details). We observed a significant inverse link between decoded uncertainty and pupil size (permutation tests, $p < 0.05$, FWER-corrected). Thus, pupil size was reliably larger when the cortical representation of the stimulus was more precise (Figure 4C, left panel). Interestingly, the timing of the effect overlapped strongly with stimulus presentation, consistent with the idea that arousal modulates the quality of the stimulus representation in cortex.

We then turned to the rate at which pupil size changed prior to and during stimulus presentation. We first estimated the slope of the pupil response in a specified time window (sliding window of length 500 ms and 1000 ms), and took this slope as a measure of change (see Methods). We then computed the across-trial correlation coefficient between pupil slope and decoded uncertainty. This revealed a significant inverse relationship between pupil slope and decoded uncertainty prior to the onset of the cue (permutation tests, $p < 0.05$, FWER-corrected; see Figure 4C, left panel and inset, 500 and 1000 ms windows), which lasted until the onset of the stimulus (Figure 4C, left, inset, 1000 ms window; see Figure 4D for individual correlation coefficients and an example observer). Thus, it appears that pupil size first dilates in anticipation of the stimulus, and then remains constant at increased size during stimulus presentation (Figure 4C left panel and inset). Taken together, our analyses show that there is a reliable relationship between decoded uncertainty and both pupil size and dilation. This altogether suggests that spontaneous fluctuations in arousal state result in an improved representation of stimulus orientation in the human visual cortex.

Relationship between pupil-linked arousal and reported confidence

The participants not only reported the presented orientation, but also gave their level of confidence in this judgment. We previously showed, using the same dataset as here, that reported levels of confidence are linked to both behavioral performance and the precision of the cortical stimulus representation. This suggests that subjective confidence is computed from the degree of uncertainty in cortex (Geurts et al., 2022). Based on this relationship, we here asked whether pupil-linked arousal also predicts confidence. In other words, we took reported confidence as an (indirect) measure of the degree of uncertainty in cortex to see if it is linked to the pupil's response. To address this question, we again divided the data for each individual observer into three bins of increasing confidence, computed the mean pupil size across all trials in each bin (separately for each time point), and combined the data across observers (see Methods). We compared pupil size between the first (lowest) and third (highest) confidence bins (Figure 4B, right panel). This analysis revealed a significant difference in pupil size between the high and low confidence bins, starting around 0.5-0.6 s after stimulus onset and lasting until about 1.5 s after stimulus offset (t-tests; all $p < 0.05$, FWER-corrected). That is, higher levels of confidence were reliably associated with greater pupil size, suggesting that arousal state affects the subjective level of confidence of the observers.

We next analyzed the data on a trial-by-trial basis. Specifically, much like before, we computed for each individual observer the correlation coefficient between reported confidence and pupil size (for each time point) or slope (computed over a specified sliding window of time), and combined the data across observers (Figure 4C, right panel). While we observed no reliable link with pupil size, there was a significant positive relationship between reported confidence and pupil slope during stimulus presentation (0.3-0.9 s after stimulus onset; $p < 0.05$, FWER-corrected). Thus, the steeper was the slope of the pupil's response, the more confident the observers were about their orientation judgments. Because stronger pupil dilation (or weaker constriction) is associated with higher levels of arousal

(Reimer et al., 2016), this altogether suggests that arousal state modulates both the cortical representation of orientation and reported confidence.

Pupil-linked arousal is linked to behavior

Is the link between arousal state and the quality of the cortical stimulus representation also reflected in behavior? We reasoned that if arousal state modulates the precision of information in cortex, it should similarly impact behavior. We tested this hypothesis as follows. We first divided, per participant and time point, all trials into ten bins of increasing pupil size or slope. We then quantified behavioral imprecision as the variance in the orientation estimation errors in each bin, and performed a multiple linear regression analysis to compute the partial correlation coefficient between pupil size or slope and behavioral variability, while controlling for interindividual differences in the mean. Our results indicate that pupil-linked arousal boosts behavioral performance, much like it improves neural precision. Specifically, we observed a reliable inverse correlation between pupil slope and behavioral variability prior to and during stimulus presentation (permutation tests, $p < 0.05$, FWER-corrected). The correlation coefficient between pupil size and behavioral variability during and immediately after stimulus presentation was also negative and significant (permutation tests, $p < 0.05$, FWER-corrected; Figure 5). In other words, orientation estimates were more precise on trials with stronger pupil dilation just before and during stimulus presentation, indicating a state of higher arousal. This shows that spontaneous fluctuations in arousal state manifest themselves not only at the neural level, but also in behavior.

Assessing the relative magnitude of the effect of arousal on uncertainty

Our findings suggest that arousal state modulates the precision of the cortical stimulus representation. However, it remains unclear how large the impact is of arousal on representational precision in cortex. The size of arousal's impact cannot be taken directly from its relationship with decoded uncertainty (i.e., from the magnitude of the obtained correlation coefficient), as decoded uncertainty reflects not

only neural but also many non-neural sources of noise, including from the MRI scanner. Similarly, pupil size is not a direct read-out of arousal state, and likely also reflects many other physiological processes. For this reason, we instead relied on an indirect approach and compared arousal's effect on uncertainty with that of stimulus orientation to acquire an understanding of their relative contribution in cortex. Behavioral accuracy and cortical activity are well known to vary across orientation stimuli, with poorer behavioral performance and reduced representational fidelity for oblique compared to cardinal orientations (Appelle, 1972; Furmanski and Engel, 2000; van Bergen et al., 2015). We also observed this oblique effect in our own data, with greater decoded uncertainty and larger behavioral variability for oblique compared to cardinal orientation stimuli (correlation between distance-to-cardinal and decoded uncertainty or behavioral variability, respectively: $\rho = 0.025$, $z = 2.95$, $p = 0.002$, and $r = 0.63$, $t(287) = 13.60$, $p < 0.001$; Figure 6A; see also Geurts et al., 2022, Extended Data Figure 2B). To assess the relative impact of pupil-linked arousal on representational imprecision, we compared the absolute effect sizes ($|\rho|$) between the two. Interestingly, we found that the impact of pupil-linked arousal on decoded uncertainty is of the same order of magnitude as that of stimulus orientation ($|\rho| = 0.025$ versus $|\rho| = 0.023$ for orientation and arousal, respectively, Figure 6B). This suggests that arousal state has a rather significant influence on representational fidelity in cortex – almost as large as that of a physical change in stimulus orientation.

DISCUSSION

Do spontaneous fluctuations in arousal state affect the quality of stimulus information contained in visual cortical activity? Here, we addressed this question by measuring the degree of uncertainty in the cortical stimulus representation using a probabilistic decoding technique, while taking pupil size as an index of arousal state. We observed that both pupil-linked arousal and decoded sensory uncertainty fluctuate over trials. Moreover, we discovered that these trial-by-trial fluctuations in arousal state are linked to the uncertainty contained in visual cortical activity. Specifically, trials of low sensory uncertainty differed from high uncertainty trials in that the pupil rapidly dilated just prior to stimulus onset, followed by sustained levels of increased pupil size during stimulus presentation, when uncertainty was low. Because rapid pupil dilation is a hallmark of a change in arousal state, this suggests that arousal affects the degree of uncertainty in the cortical representation of the stimulus. Interestingly, we observed a similar relationship between pupil size and subjective confidence, a secondary measure of sensory uncertainty, and between the pupil's signals and behavioral performance. A comparison between the effects of pupil-linked arousal and those of stimulus orientation suggested that arousal's impact on representational imprecision was almost as large as that of a physical change in the stimulus. Taken together, these results suggest that arousal state has reasonably large impact on the fidelity of information processing in the human visual cortex, with clear consequences for behavior.

A key distinguishing aspect of this study is that we measured the precision of the neural representation directly in cortex, using a probabilistic decoding technique that enabled us to quantify representational imprecision as the width of a probability distribution over possible stimuli. Previous studies using this technique have shown that this imprecision in the cortical stimulus representation varies from trial to trial, even when the stimulus is held constant (van Bergen et al., 2015; Geurts et al., 2022; Chetverikov and Jehee, 2023). Consistent with Bayesian theories of decision-making, these changes in imprecision have moreover been shown to affect the observer's decision-making, with larger uncertainty resulting

511 in enhanced perceptual biases (van Bergen et al., 2015; van Bergen and Jehee, 2019), lower reported
512 confidence (Geurts et al., 2022), and different perceptual choices (Walker et al., 2020). It remained
513 unclear, however, what drives such stimulus-independent fluctuations in sensory cortical uncertainty.
514 The present work builds on and extends this line of research, suggesting that arousal is one of the
515 factors influencing the imprecision in neural representations.

516 Contrary to previous studies investigating the effect of arousal on neural activity in humans (e.g. Keil
517 et al., 2003; Warren et al., 2016; Gelbard-Sagiv et al., 2018), we did not explicitly manipulate arousal,
518 but specifically focused on spontaneous fluctuations. That is, we were interested in what drives
519 variability in the precision of neural representations in the absence of external, experimentally
520 manipulated change. We used pupil size as an index of arousal, because of its well-established links to
521 the neuromodulatory systems underlying arousal (Aston-Jones and Cohen, 2005; Murphy et al., 2014;
522 Varazzani et al., 2015; Joshi et al., 2016). Interestingly, our findings are consistent with earlier work in
523 which catecholamine (noradrenaline and dopamine) levels were manipulated pharmacologically and
524 representational precision was measured across trials (Warren et al., 2016). Here we show that the
525 relationship between arousal and representational precision similarly holds on a trial-by-trial basis and
526 in the absence of explicit manipulations of arousal or noradrenaline levels.

527 What are the neural mechanisms by which arousal could modulate the precision of the stimulus
528 representation in the human visual cortex? Behavioral studies in humans have reported greater
529 contrast sensitivity with an increase in arousal (Lee et al., 2014; Kim et al., 2017), possibly mediated by
530 multiplicative effects on the underlying cortical response (Kim et al., 2017). Neurophysiological studies
531 in mice and rabbits have shown that an increase in arousal results in enhanced, more selective and
532 reliable responses to visual stimuli, and weaker noise correlations (Cano et al., 2006; Niell and Stryker,
533 2010; Erisken et al., 2014; Reimer et al., 2014; Vinck et al., 2015) – mechanisms that could all lead to
534 an increase in the amount of information contained in neural activity. Theoretical work has linked
535 arousal to changes in neural response gain (Servan-Schreiber et al., 1990; Aston-Jones and Cohen,

2005), which could similarly improve the quality of the information encoded in neural activity and, as such, reduce sensory uncertainty (Seung and Sompolinsky, 1993; Ma et al., 2006). Indeed, one neurophysiological study in monkeys directly related spontaneous fluctuations in activity (as we relied on here) to changes in neural excitability or gain (Goris et al., 2014). It altogether seems plausible that one of these mechanisms, or a combination thereof, could mediate arousal-linked fluctuations in sensory uncertainty in the human visual cortex.

It is important to realize that we do not intend to argue that arousal is the sole driver of spontaneous fluctuations in cortical information. For example, it is well known that attending to a visual feature or location improves its representation in cortex (Kamitani and Tong, 2005; Saproo and Serences, 2010; Jehee et al., 2011). Attention has also been shown to modulate pupil size (see Strauch et al., 2022 for a review). It is highly conceivable that also attention-based processes spontaneously wax and wane and affect the amount of information in cortex, much like we observed for arousal here. One way to distinguish between these processes could be to focus on the different neural systems that mediate their effects, such as the LC for arousal (Moruzzi and Magoun, 1949; Berridge and Waterhouse, 2003; Aston-Jones and Cohen, 2005; Sara and Bouret, 2012). However, the current study was not designed nor optimized for this research question. Nevertheless, a preliminary analysis showed hints of a link between LC activity and decoded uncertainty in our dataset, which further supports the notion that changes in arousal state underlie the observed uncertainty fluctuations. It will be interesting for future studies to further investigate and disentangle these and other cognitive processes that give rise to spontaneous fluctuations in neural information.

Taken together, we showed that spontaneous, trial-by-trial fluctuations in arousal state, as indexed by pupil size, are linked to the quality of visual cortical stimulus representations, as well as reported levels of subjective confidence and behavioral performance. This suggests that arousal is one of the driving

559 factors of variability in neural responses and the precision with which sensory information is encoded
560 in cortical activity.

561 REFERENCES

562 Appelle S (1972) Perception and discrimination as a function of stimulus orientation: the "oblique
563 effect" in man and animals. Psychol Bull 78:266–278.

564 Aston-Jones G, Cohen JD (2005) Adaptive gain and the role of the locus coeruleus–norepinephrine
565 system in optimal performance. J Comp Neurol 493:99–110.

566 Berridge CW, Waterhouse BD (2003) The locus coeruleus-noradrenergic system: modulation of
567 behavioral state and state-dependent cognitive processes. Brain Res Brain Res Rev 42:33–84.

568 Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

569 Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual
570 cortex. J Neurosci 29:13992–14003.

571 Campbell FW, Green DG (1965) Optical and retinal factors affecting visual resolution. J Physiol
572 181:576–593.

573 Campbell FW, Gregory AH (1960) Effect of size of pupil on visual acuity. Nature 187:1121–1123.

574 Cano M, Bezdudnaya T, Swadlow HA, Alonso J-M (2006) Brain state and contrast sensitivity in the
575 awake visual thalamus. Nat Neurosci 9:1240–1242.

576 Chetverikov A, Jehee JFM (2023) Motion direction is represented as a bimodal probability distribution
577 in the human visual cortex. Nat Commun 14:7634.

578 Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface
579 reconstruction. NeuroImage 9:179–194.

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

582 de Gee JW, Tsetsos K, Schwabe L, Urai AE, McCormick D, McGinley MJ, Donner TH (2020) Pupil-linked
583 phasic arousal predicts a reduction of choice bias across species and decision domains. eLife
584 9:1–25.

585 Deyoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996) Mapping striate
586 and extrastriate visual areas in human cerebral cortex. Proc Natl Acad Sci U S A 93:2382–2386.

587 Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial
588 precision of functional MRI. Cereb Cortex 7:181–192.

589 Erisken S, Vaiceliunaite A, Jurjut O, Fiorini M, Katzner S, Busse L (2014) Effects of Locomotion Extend
590 throughout the Mouse Early Visual System. Curr Biol 24:2899–2907.

591 Faisal AA, Selen LPJ, Wolpert DM (2008) Noise in the nervous system. Nat Rev Neurosci 9:292–303.

592 Fieller EC, Pearson ES (1961) Tests for Rank Correlation Coefficients: II. Biometrika 48:29.

593 Furmanski CS, Engel SA (2000) An oblique effect in human primary visual cortex. Nat Neurosci 3:535–
594 536.

595 Gelbard-Sagiv H, Magidov E, Sharon H, Hendler T, Nir Y (2018) Noradrenaline Modulates Visual
596 Perception and Late Visually Evoked Activity. Curr Biol 28:2239-2249.e6.

597 Geurts LS, Cooke JRH, van Bergen RS, Jehee JFM (2022) Subjective confidence reflects representation
598 of Bayesian probability in cortex. Nat Hum Behav 6:294–305.

599 Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD (2010) Pupil diameter tracks changes in control state
600 predicted by the adaptive gain theory of locus coeruleus function. Cogn Affect Behav Neurosci
601 10:252–269.

602 Goris RLT, Movshon JA, Simoncelli EP (2014) Partitioning neuronal variability. Nat Neurosci 17:858–
603 865.

604 Hedges LV, Olkin I (1985) Statistical methods for meta-analysis. Academic Press.

605 Jehee JFM, Brady DK, Tong F (2011) Attention Improves Encoding of Task-Relevant Features in the
606 Human Visual Cortex. J Neurosci 31:8210–8219.

607 Jenkinson M, Bannister P, Brady M, Smith SM (2002) Improved optimization for the robust and
608 accurate linear registration and motion correction of brain images. NeuroImage 17:825–841.

609 Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. NeuroImage 62:782–
610 790.

611 Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain
612 images. Med Image Anal 5:143–156.

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

615 Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat
616 Neurosci 8:679–685.

617 Keil A, Gruber T, Müller MM, Moratti S, Stolarova M, Bradley MM, Lang PJ (2003) Early modulation of
618 visual perception by emotional arousal: Evidence from steady-state visual evoked brain
619 potentials. Cogn Affect Behav Neurosci 3:195–206.

620 Kim D, Lokey S, Ling S (2017) Elevated arousal levels enhance contrast perception. J Vis 17:14.

621 Kleiner M, Brainard DH, Pelli DG (2007) What's new in Psychtoolbox-3? Perception 36.

622 Knapen T, de Gee JW, Brascamp J, Nuiten S, Hoppenbrouwers S, Theeuwes J (2016) Cognitive and
623 Ocular Factors Jointly Determine Pupil Responses under Equiluminance. PLOS ONE
624 11:e0155574.

625 Lee T-H, Baek J, Lu Z-L, Mather M (2014) How arousal modulates the visual contrast sensitivity function.
626 Emotion 14:978–984.

Li H-H, Sprague TC, Yoo AH, Ma WJ, Curtis CE (2021) Joint representation of working memory and uncertainty in human cortex. Neuron.

Livingstone MS, Hubel DH (1981) Effects of sleep and arousal on the processing of visual information in the cat. Nature 291:554–561.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nat Neurosci 9:1432–1438.

Mathôt S (2018) Pupillometry: Psychology, Physiology, and Function. J Cogn 1:1–23.

Moruzzi G, Magoun HW (1949) Brain stem reticular formation and activation of the EEG. Electroencephalogr Clin Neurophysiol 1:455–473.

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

Nichols TE, Hayasaka S (2003) Controlling the familywise error rate in functional neuroimaging: A comparative review. Stat Methods Med Res 12:419–446.

Niell CM, Stryker MP (2010) Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. Neuron 65:472–479.

Pfeffer T, Keitel C, Kluger DS, Keitel A, Russmann A, Thut G, Donner TH, Gross J (2022) Coupling of pupil- and neuronal population dynamics reveals diverse influences of arousal on cortical processing. eLife 11:71890.

Podvalny E, King LE, He BJ (2021) Spectral signature and behavioral consequence of spontaneous shifts of pupil-linked arousal in human. eLife 10:e68265.

Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, Tolias AS (2014) Pupil Fluctuations Track Fast Switching of Cortical States during Quiet Wakefulness. Neuron 84:355–362.

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 7:1–7.

Reuter M, Schmansky NJ, Rosas HD, Fischl B (2012) Within-subject template estimation for unbiased longitudinal image analysis. NeuroImage 61:1402–1418.

Roth ZN, Ryoo M, Merriam EP (2020) Task-related activity in human visual cortex. PLOS Biol 18:e3000921.

Saproo S, Serences JT (2010) Spatial Attention Improves the Quality of Population Codes in Human Visual Cortex. J Neurophysiol 104:885–895.

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

Sara SJ, Bouret S (2012) Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal. Neuron 76:130–141.

Schröder S, Steinmetz NA, Krumin M, Pachitariu M, Rizzi M, Lagnado L, Harris KD, Carandini M (2020) Arousal Modulates Retinal Output. Neuron 107:487-495.e9.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Servan-Schreiber D, Printz H, Cohen JD (1990) A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. Science 249:892–895.

Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. Proc Natl Acad Sci 90:10749–10753.

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44:83–98.

Strauch C, Wang CA, Einhäuser W, Van der Stigchel S, Naber M (2022) Pupillometry as an integrated readout of distinct attentional networks. Trends Neurosci 45:635–647.

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

van Bergen RS, Jehee JFM (2018) Modeling correlated noise is necessary to decode uncertainty. NeuroImage 180:78–87.

van Bergen RS, Jehee JFM (2019) Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. J Neurosci 39:8164–8176.

van Bergen RS, Ma WJ, Pratte MS, Jehee JFM (2015) Sensory uncertainty decoded from visual cortex predicts behavior. Nat Neurosci 1:1–6.

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 35:7866–7877.

Vinck M, Batista-Brito R, Knoblich U, Cardin JA (2015) Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. Neuron 86:740–754.

Walker EY, Cotton RJ, Ma WJ, Tolias AS (2020) A neural basis of probabilistic computation in visual cortex. Nat Neurosci 23:122–129.

Warren CM, Eldar E, van den Brink RL, Tona KD, van der Wee NJ, Giltay EJ, Van Noorden MS, Bosch JA, Wilson RC, Cohen JD, Nieuwenhuis S (2016) Catecholamine-mediated increases in gain enhance the precision of cortical representations. J Neurosci 36:5699–5708.

FIGURE LEGENDS

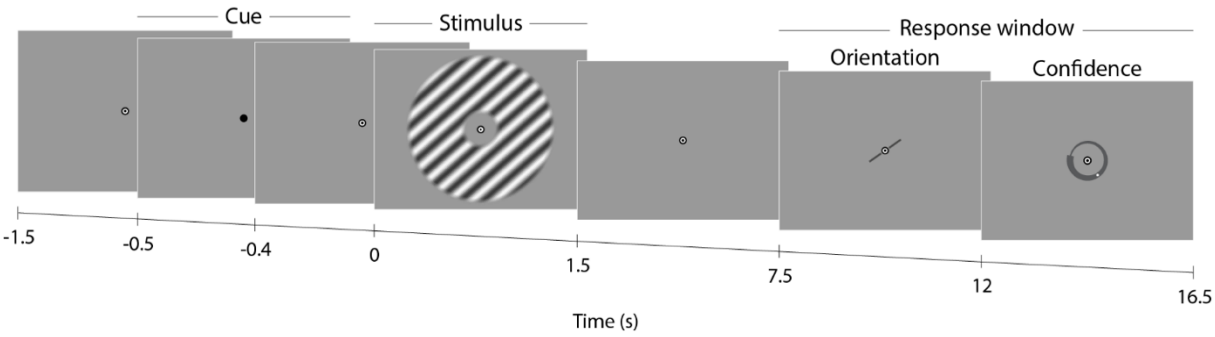

Figure 1: Overview of the orientation estimation task. Participants were required to fixate the bullseye target in the center of the screen throughout each run. Each trial started with this bullseye briefly turning black, as a cue to stimulus onset. The visual stimulus was a counterphasing sinusoidal grating, presented in an annulus around fixation, and was presented for 1.5 s. After a 6 s-delay (fixation interval), a black bar appeared in the center of the screen, and participants were required to report their orientation estimate by rotating this bar. Next, they reported their level of confidence about this estimate by sliding a dot over the (circular) confidence scale. Both response intervals lasted for 4.5 s, and the bar or scale started fading after 3.5 s to indicating the approaching end of the response window. A 1.5-s intertrial interval separated the response window from the next stimulus. Participants completed 20 trials per run.

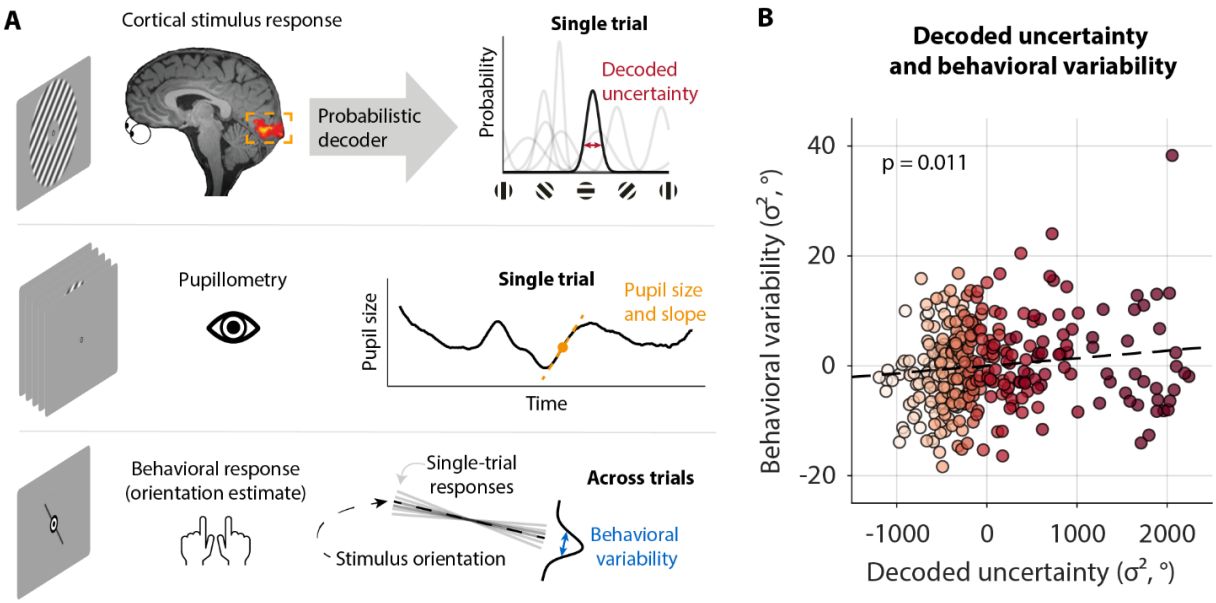

Figure 2: Overview of behavioral, physiological, and neural measures and behavioral benchmarking. A) The imprecision in the cortical stimulus representation was quantified using a probabilistic decoding technique. This decoding algorithm computes a probability distribution over stimulus orientation from single-trial activity patterns. The width of the distribution was taken as a measure of the degree of uncertainty in the cortical stimulus representation. The observer's perceptual uncertainty was quantified by the width of the distribution of behavioral responses across trials. Pupil size and the slope (the first temporal derivative) of the pupil signal were used to measure arousal over the course of each trial. Behavioral imprecision was measured as the variability in the observer's orientation judgments across trials. B) Decoded uncertainty predicts behavioral variability on a trial-by-trial basis. Each observer's trials were split into ten bins of increasing decoded uncertainty. Behavioral variability was computed as the squared circular standard deviation of behavioral response errors in each bin. The partial correlation coefficient between decoded uncertainty and behavioral variability (controlling for difference in intercept between participants) was computed and found to be significantly positive ($t(287) = 2.30$, $p = 0.011$, $r = 0.13$, 95% CI = [0.019, 0.25]). Thus, when the cortical stimulus representation was more uncertain, behavioral imprecision was larger, as well. Note that the data is centered around zero because this is a partial correlation plot so interindividual differences in the

724 mean have been removed. The fMRI results in (B) were also reported in Geurts et al. (2022), Extended
725 Data Fig. 2.

726

727

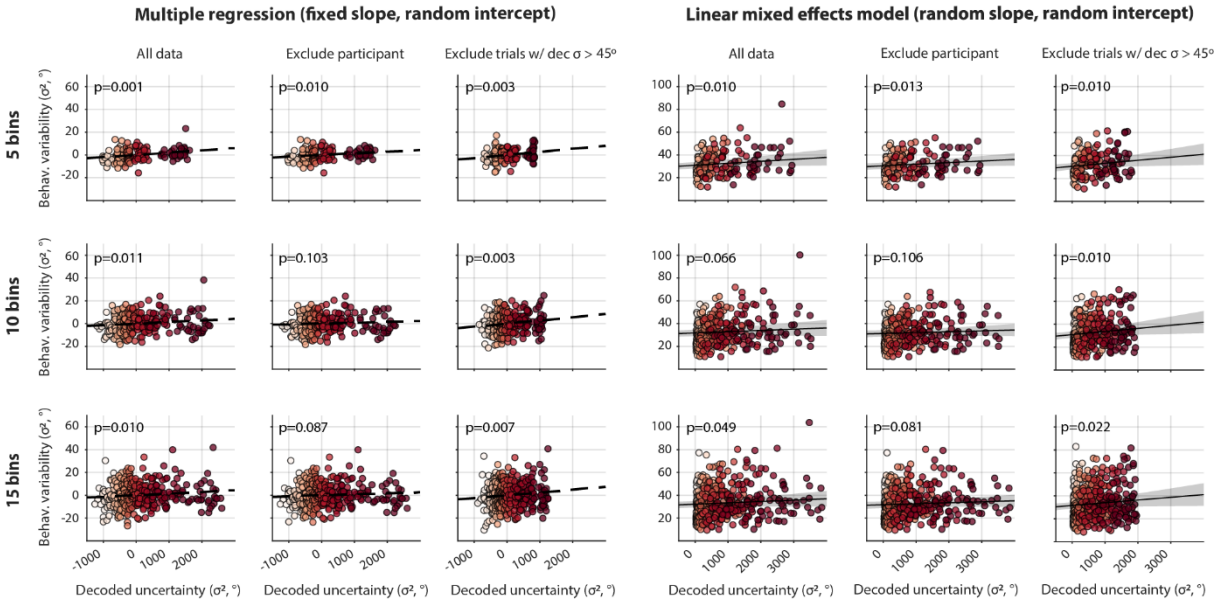

Figure 3: Control analyses for the relationship between decoded uncertainty and behavioral variability. Three parameters were varied in the analyses: 1) The number of uncertainty bins; that is, five, ten or fifteen bins per participant. 2) The statistical model; specifically, we modeled the strength (slope) of the effect as a fixed effect (multiple regression) or as a random effect (linear mixed-effects model). 3) The exclusion criteria; specifically, we either excluded the observer who gave rise to the extreme data point in the top-right corner, or excluded all trials on which the decoded level of uncertainty was unrealistically high (S.D. of the decoded distribution > 45 degrees). The reasoning behind the latter approach is that such very high values of decoded uncertainty (corresponding to a very wide or almost flat decoded distribution) likely reflect non-neural sources of noise related to, for example, the MRI scanner. We found that our results are fairly robust to these changes in analysis parameters. In fact, we even observed a *stronger* relationship between decoded uncertainty and behavioral variability when trials with unrealistically high decoded uncertainty were removed altogether. This shows that the observed positive correlation coefficient between decoded uncertainty and behavioral variability is not driven by extreme values, and strengthens our conclusion that decoded uncertainty reflects the precision of the underlying neural representation. Please note that this replicates previous findings (van Bergen et al., 2015; Chetverikov and Jehee, 2023).

Data points represent individual observers and colors indicate bin numbers (1-5, 1-10, or 1-15 depending on the number of bins used). Dashed lines indicate the best (linear) fit, shaded area represents 95% confidence interval. For multiple regression, partial residuals are shown, which is why the data are centered around zero.

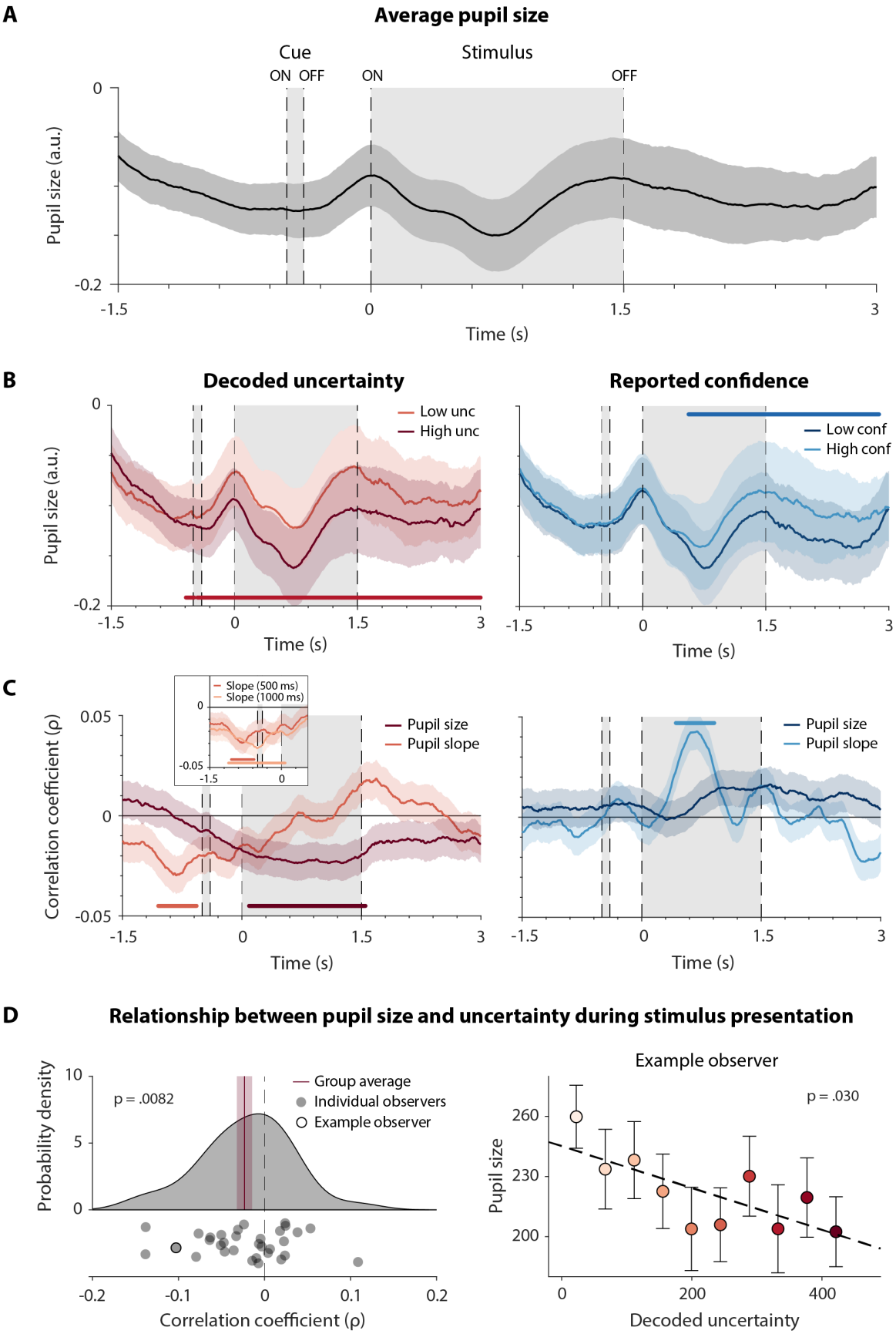

Figure 4: Relationship between arousal state and the degree of uncertainty in the cortical stimulus

representation. A) Mean pupil size over time. The cue appeared at -0.5 s and lasted for 0.1 s, and the stimulus was on screen from 0 to 1.5 s (0 is when the stimulus appeared on the screen). B) Pupil size timeseries separated for high and low decoded uncertainty (left) or reported confidence (right). Trials were divided into three bins of increasing uncertainty (confidence); shown are the first and third bin. C) Trial-by-trial Spearman correlation coefficients between pupil size or slope (500 ms sliding window) and decoded uncertainty (left) or reported confidence (right). Inset in left panel shows correlation between decoded uncertainty and pupil slope for two sliding windows of different length (500 and 1000 ms). A-C) Shaded areas represent the standard error of the mean (s.e.m.) across observers. Bars indicate significance (permutation tests, $p < 0.05$, FWER-corrected). D) Relationship between decoded uncertainty and pupil size during stimulus presentation. Pupil size was averaged over the stimulus presentation window to obtain a single value per trial. Left: Distribution of Spearman correlation coefficients for individual observers, and group average (shaded area indicates s.e.m.). Right: Relationship between pupil size and decoded uncertainty for an example observer. Decoded uncertainty and pupil size were ranked, trials were subsequently divided into ten bins based on decoded uncertainty, and the data was averaged within each bin. Error bars represent s.e.m. across trials.

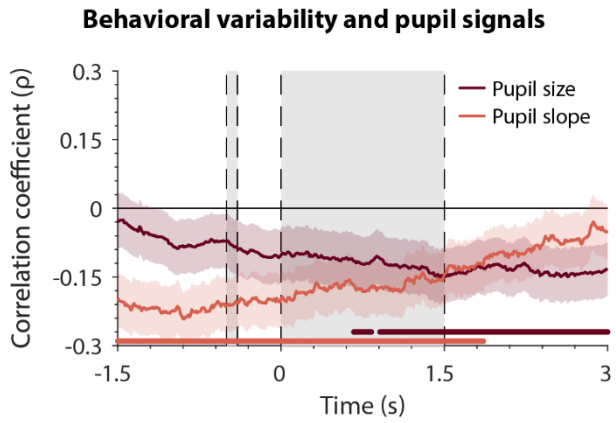

Figure 5: Behavioral imprecision is linked to arousal state. The correlation between pupil size or slope and behavioral variability was calculated for each point in time, in a window from 1.5 s before stimulus onset until 1.5 s after stimulus offset. For each time point and participant, trials were divided into ten bins of increasing pupil size (slope), and the variance in behavioral orientation judgments was computed per bin. The correlation coefficient between pupil measures and behavioral variability was computed while controlling for differences in mean (intercept), and is plotted over time. We observed that increased pupil-linked arousal – as indicated by both pupil slope and size – results in more precise (less variable) behavioral responses. Shaded areas represent the standard error of the partial correlation coefficients. Bars indicate significance (permutation tests, $p < 0.05$, FWER-corrected). Dashed lines indicate on- and offsets of the cue and stimulus.

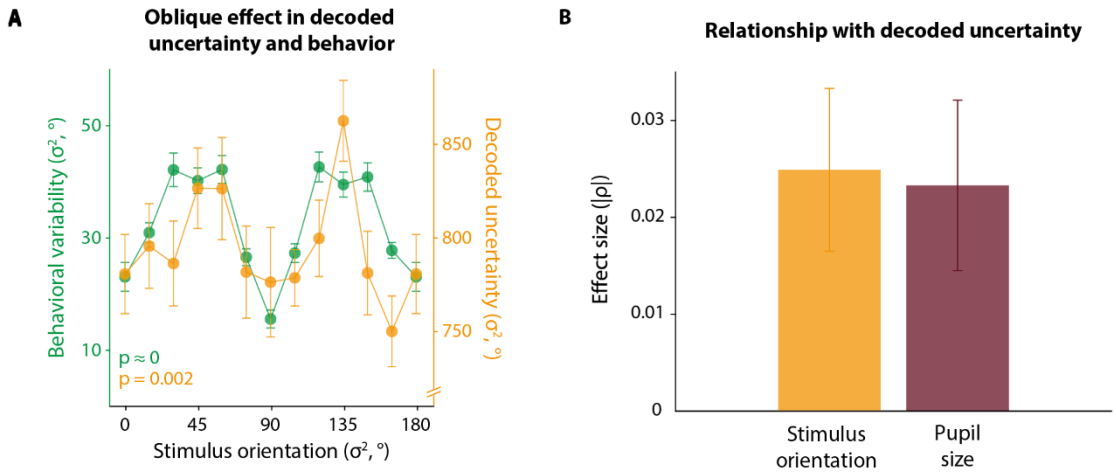

Figure 6: Effect sizes of arousal state and stimulus orientation are comparable. A) Decoded uncertainty is greater for oblique compared to cardinal orientation stimuli (correlation between distance-to-cardinal and decoded uncertainty: $\rho = 0.025$, $z = 2.95$, $p = 0.002$). This finding is paralleled by imprecision in the observer's behavior (correlation between distance-to-cardinal and behavioral variability: $r = 0.63$, $t(287) = 13.60$, $p < 0.001$). B) The strength of uncertainty's relationship with stimulus orientation, $|\rho| = 0.025$, is comparable to that with pupil-linked arousal, $|\rho| = 0.023$. Pupil size was averaged over the stimulus presentation window to obtain a single value per trial, and then correlated with trial-by-trial decoded uncertainty. Panel (A) was also reported in Geurts et al. (2022), Extended Data Fig. 2, and is reproduced here for convenience.