Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating

Alexandre L.S. Filipowicz[1,2,3]*, Christopher M. Glaze[1,2,3]*, Joseph W. Kable[2,3] & Joshua I. Gold[1,3]

Departments of [1]Neursocience and [2]Psychology, [3]Computational Neuroscience Initiative, University of Pennsylvania
*co-first authors

Corresponding author:
Alexandre Filipowicz
Richards Medical Research Building
3700 Hamilton Walk, D503
University of Pennsylvania
Philadelphia, PA
19104, USA

alsfilip@pennmedicine.upenn.edu

Pupils tend to dilate in response to surprising events, but whether these responses are primarily stimulus driven or instead reflect a more nuanced relationship between pupil-linked arousal systems and cognitive expectations is not known. Using an auditory adaptive decision-making task, we show that evoked pupil diameter is modulated more strongly by violations of learned, top-down expectations than by changes in low-level stimulus properties. We further show that both baseline and evoked pupil diameter is modulated by the degree to which individual subjects use these violations to update their subsequent expectations, as reflected in the complexity of their updating strategy. Together these results demonstrate a central role for cognitive processing in how arousal systems respond to new inputs and, via our complexity-based analyses, provide a unified framework for understanding these effects in terms of both inference processes aimed to reduce belief uncertainty and more traditional notions of mental effort.

Keywords: **pupil diameter, uncertainty, surprise, complexity, adaptive decision-making, mental models**

Introduction

Pupil diameter changes largely as a function of ambient light levels but also in the context of a host of other brain functions that involve changes in arousal (Mathôt, 2018). Several early studies ascribed arousal-related modulations of pupil diameter to the "mental effort" required to, for example, solve an arithmetic problem or hold items in working memory (Beatty, 1982; Mathôt, 2018). This view emphasizes the idiosyncratic and cognitive nature of these pupil fluctuations but lacks a precise description of the computations being encoded. More recently, arousal-related pupil modulations have been shown to represent critical computations required to learn from, and adapt to, changes in our environment (Krishnamurthy, Nassar, Sarode, & Gold, 2017; Nassar et al., 2012; O'Reilly et al., 2013; Preuschoff, 't Hart, & Einhäuser, 2011). This view emphasizes specific computations encoded in pupil diameter but lacks a clear link to idiosyncratic, cognitive processing. The goal of the present study was to reconcile these different views and synthesize a more comprehensive understanding of the role of pupil-linked arousal systems in cognition.

We build on studies of adaptive decision-making, in which past experiences are used to build and adjust models of the world that guide decisions (Filipowicz, Anderson, & Danckert, 2016; Johnson-Laird, 2004; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). In uncertain and changing environments, effective models need to resist spurious noise while remaining flexible enough to adapt to real changes (Behrens, Woolrich, Walton, & Rushworth, 2007; Filipowicz et al., 2016; Nassar, Wilson, Heasly, & Gold, 2010; O'Reilly, 2013). Several key features of these adaptive processes are reflected robustly in dynamic pupil fluctuations. Rapid phasic increases in pupil diameter occur in response to surprising events that can drive changes in the model used by the decision-maker (Nassar et al., 2012; O'Reilly et al., 2013; Preuschoff et al., 2011). Slower tonic increases in pupil diameter can encode increased model uncertainty that can be reduced with additional information (Krishnamurthy et al., 2017; Muller, Mars, Behrens, & Reilly, 2019; Nassar et al., 2012; Urai, Braun, & Donner, 2017). These slower pupil changes are also those most closely linked with changes in mental effort (Beatty, 1982; Mathôt, 2018) and associated with individual differences in cognitive traits such as fluid intelligence (Tsukahara, Harrison, & Engle, 2016).

However, several key cognitive and idiosyncratic aspects of adaptive decision-making have yet to be linked clearly to pupil modulations. First, adaptive decision-making has been studied primarily in relatively stable environments, in which changes to stimulus properties occur infrequently and are therefore almost always surprising (Filipowicz et al., 2016; Krishnamurthy et al., 2017; Nassar et al., 2012; O'Reilly et al., 2013). These environments can confound top-down, cognitive expectations that govern uncertainty with bottom-up sensory responses to stimulus changes that can also modulate pupil diameter (e.g., "startle" or "orienting" responses; Bremner, 2009; Sara, 2009)). Second, adaptive decision-making requires identifying changes in properties of not just observable stimuli but also of latent (i.e., not directly observable) environmental variables that affect uncertainty (Glaze, Filipowicz, Kable, Balasubramanian, & Gold, 2018; Glaze, Kable, & Gold, 2015; Krishnamurthy et al., 2017; McGuire, Nassar, Gold, & Kable, 2014). Third, computations related to surprise and uncertainty can be strongly idiosyncratic, reflecting the specific mental model being used by the decision-maker to integrate information from the environment (Faraji, 2018; Filipowicz, Valadao, Anderson, & Danckert, 2014, 2018).

The current study aimed to overcome these limitations and assess pupil diameter in a range of cognitively demanding task conditions and with respect to individual differences in the mental models used to solve the task (Glaze et al., 2018). Using a task designed to dissociate stimulus

changes from belief updating, we show that phasic and tonic pupil signals encode surprise and uncertainty, respectively, in an adaptive, context-dependent manner. Additionally, these pupil modulations depend on the idiosyncratic decision strategies used by different subjects, which we characterized in terms of the "complexity" of the underlying mental model and which provides a quantitative description of the specific cognitive demands that evoke changes in pupil-linked arousal.

Results

Seventy-eight subjects performed a predictive-inference task in which they predicted on each trial which of two sources (left or right) would generate an auditory tone (Fig. 1a). The source generating the tones switched at unannounced points throughout the task. Critically, the probability of a source switch, or 'hazard rate' (H), also changed at unannounced points, such that during certain parts of the task source switches occurred at low (H=0.01), intermediate (H=0.30), or high (H=0.99) rates (Fig. 1b). These task conditions allowed us to dissociate expectations from stimulus properties: the low versus high hazard-rate conditions produced stimulus switches that were either unexpected or expected, respectively, whereas the intermediate hazard-rate condition minimized expectations, such that the best strategy was to match responses to the tone heard on the previous trial (similar to H=0.5 for a perceptual-judgment task; Glaze et al., 2018, 2015). Thus, inferring the current hazard rate could improve predictions about the subsequent source location. To make this inference process more challenging, uncertainty was added by playing the tone from the opposite side of the generating source on 20% of trials.

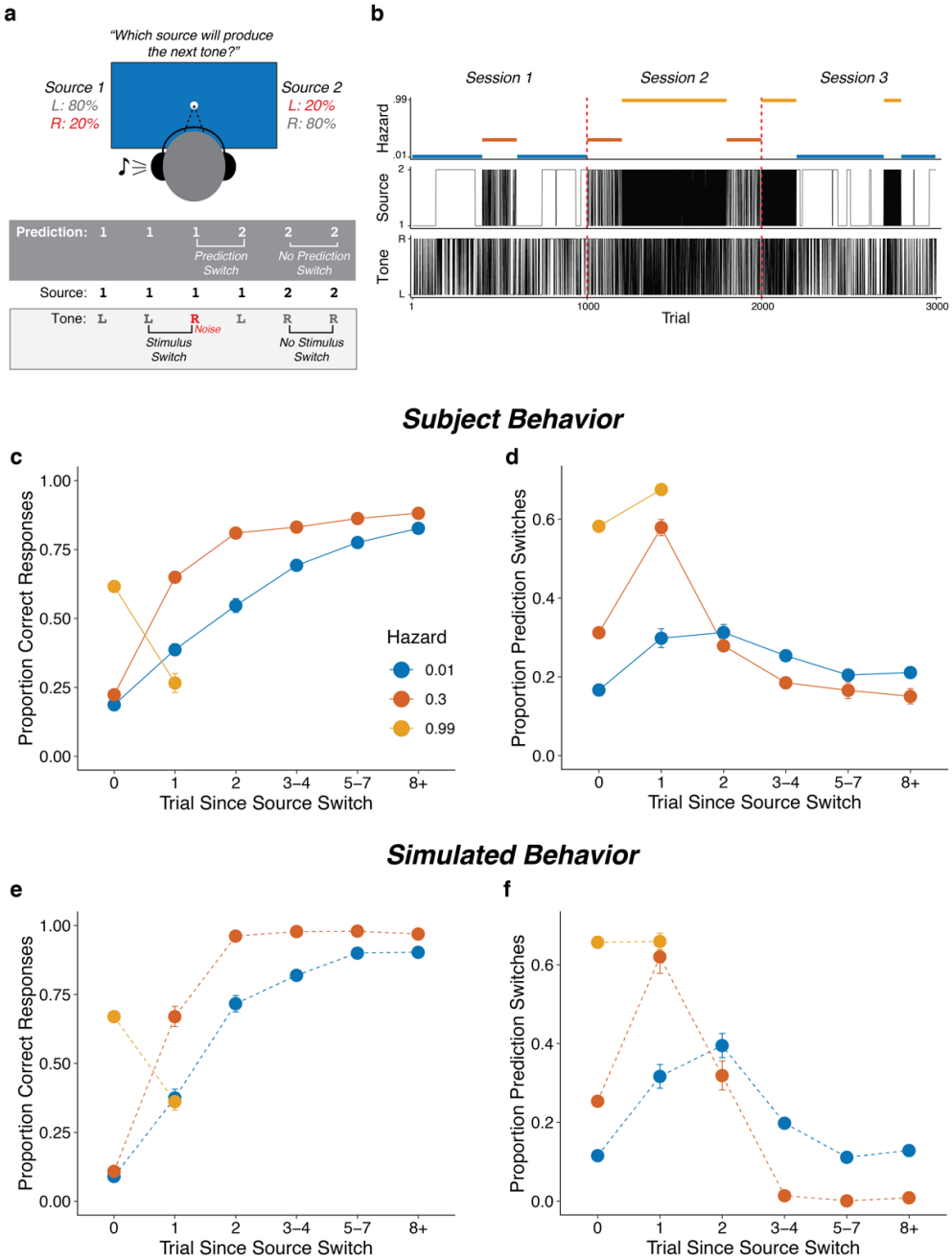

Figure 1. Auditory prediction task. a) Subjects predicted which of two sources would generate a tone on the next trial. After making the prediction, a tone was played in either the left or right ear. b) Subjects completed up to three sessions of 1000 trials each. In each session, subjects were exposed to two hazard rates that would switch at unannounced points throughout the session. c) Subject proportion correct responses as a function of trials after a source switch, separated by hazard rate. d) Proportion of subject prediction switches (i.e., changing their prediction from the previous trial) for trials after a source switch, separated by hazard rate. e,f) same as c and d but using simulated data from the hazard learning model fit to data from individual subjects and applied to the same sequences observed by each subject. In c–f, error bars are ±SEM across real (c,d) or simulated (e,f) subjects.

Pupil diameter is sensitive to adaptive, context-dependent expectations

On average, subjects adjusted their behavior adaptively to different hazard rates, in a manner that was consistent with behavioral results from previous studies and with an adaptive, Bayesian learning model (Glaze et al., 2018, 2015; Fig. 1c-f). On low hazard-rate trials, changes were unexpected. Accordingly, performance was low on trials in which the source switched but then gradually increased over several trials as the subject gathered more evidence that a real source switch had occurred. On intermediate hazard-rate trials, changes were more expected, which was reflected in a higher propensity of the subjects to switch responses after a real source switch and then show a steeper improvement on subsequent trials. On high hazard-rate trials, changes were highly expected, which was reflected in a relatively high prediction switch rate on source-switch trials, with more errors on the rare non-switch trials. These hazard-dependent behavioral patterns were consistent with simulated behavior from a Bayesian learning model that uses trial-by-trial observations to perform inference over both the current source and hazard rate (Fig. 1e,f; Glaze et al., 2018). Overall, performance was highest in the low hazard-rate condition followed by the high hazard-rate condition (median [interquartile range] proportion correct responses for low, intermediate, and high hazard rates respectively: .83 [.78-.85], .59 [.57-.61], and .63 [.56-.67]; paired rank-sum test: low vs. intermediate, $p = 2.9\text{x}10^{-9}$; low vs. high, $p = 5.4\text{x}10^{-10}$; intermediate vs. high, $p = .0009$), but above chance in all three conditions (sign-rank test for H_0: median=0.5: all ps $< 3.7\text{x}10^{-11}$).

We measured pupil diameter before ("baseline") and during ("evoked") stimulus presentation on each trial. Overall, baseline pupils tended to be largest, and evoked changes smallest, in the low hazard-rate condition (linear mixed effects contrast between low versus other hazard conditions: all baseline ps<0.032, all evoked-change ps<0.0007). Baseline and evoked changes did not differ between the intermediate and high hazard-rate conditions (all ps>0.470). We next focused our analyses on the degree to which these hazard-specific pupil measurements were modulated by whether or not there was a switch (i.e., from left to right or right to left) in either the subject's prediction or the actual stimulus location on a given trial, which could be distinguished from each other because they did not always occur on the same trials (Fig. 1a).

Both baseline and evoked pupil diameter encoded prediction and stimulus switching in a manner that reflected the top-down, hazard-specific expectations that governed task performance (Fig. 2). On low and intermediate hazard-rate trials, in which prediction and stimulus switches

were unexpected, the baseline pupil diameter just preceding the prediction was larger on trials with prediction switches but not the yet-to-occur stimulus switches. After tone onset, evoked pupil changes were larger for both prediction and stimulus switches. These effects persisted into the next baseline pupil, with larger pupil dilation after both stimulus and prediction switches on low and intermediate hazard-rate trials. In contrast, on high hazard-rate trials, in which switches were expected, prediction and stimulus switches resulted in pupil constrictions, although the influence of stimulus switches on evoked pupil changes was less evident than in the low or intermediate hazard conditions. These effects also persisted into the next trial (Fig. 2c). Thus, pupil dilations occurred in response to low-level stimulus properties (stimulus switches) but more generally to violations of learned, hazard-dependent expectations, corresponding to switches when stability was expected and repetitions when instability was expected.

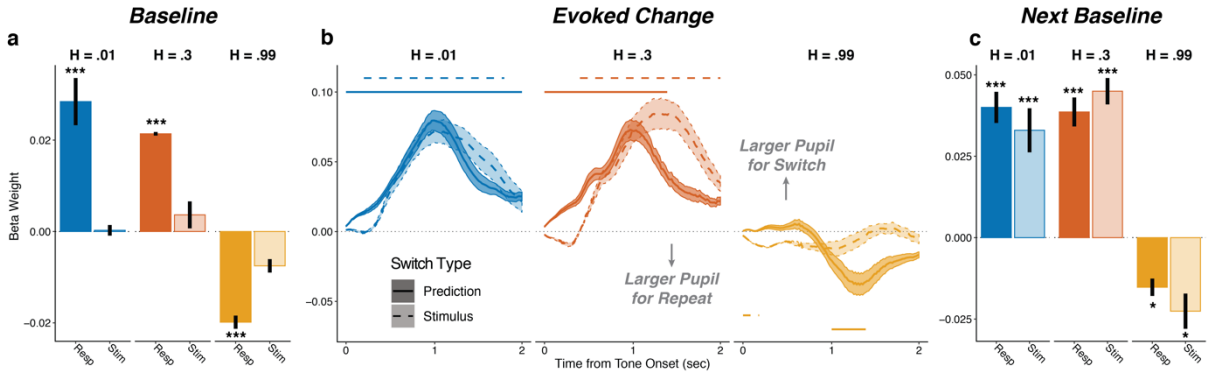

Figure 2. Pupil encodes violations of expectations. a) Linear mixed effects beta weights for baseline (100 ms prior to tone onset) pupil dynamics related to whether subjects switched their prediction after hearing the tone (Pred) or whether the tone stimulus switched sides on that trial (Stim). b) Beta weights for prediction and stimulus switches, as indicated, over a 2 s period after tone onset. c) Beta weights for the influence of prediction and stimulus switches on the next trial's baseline pupil. Both baseline pupil measures accounted for the baseline pupil and evoked pupil change on the previous trial. The evoked change and next baseline pupil measures accounted for the previous baseline pupil and the evoked pupil change and baseline pupil on the current trial. Bars and errorbars in a and c or lines and shading in b correspond to mean±SEM across subjects. Straight lines above and below beta values in panel b indicate significant timepoints (FWE corrected $p<0.05$). In bar plots, *: $p<0.05$, ***: $p<0.001$.

Pupil diameter and response times encode adaptive, trial-by-trial belief updating

To better relate the task-dependent pupil modulations to the specific updating strategies each individual subject used to derive hazard-specific predictions, each subject's trial-by-trial sequence of behavioral choices (predictions) were fit with a Bayesian learning model that performs inference over both the source location and hazard rate (Glaze et al., 2018). A key parameter in the model controls the width of the distribution of prior probabilities that subjects assign to different hazard rates, with wider prior distributions leading to model predictions that adapt more effectively to changes in hazard rate than narrow prior distribution (Fig. S1a). The model uses this prior distribution along with each new observation (i.e., on which side the sound is played) to update its strength of belief about the current source location (quantified as a log-odds-ratio that the left or right source would generate the next tone; Gold & Shadlen, 2001) and hazard rate. Surprise (measured as the negative log probability that the observed tone was predicted) reflects the degree to which the new observation does not match the current belief about location (Filipowicz et al., 2018; Mars et al., 2008; O'Reilly et al., 2013; Fig. S1b).

Baseline and evoked pupil diameter encoded belief strength and surprise, respectively, from this belief-updating process (Fig. 3). This encoding was similar for the three hazard-rate conditions, implying that the hazard-dependent effects of prediction and stimulus switching on pupil depicted in Fig. 2 reflected a common underlying relationship between pupil and belief updating. In particular, baseline pupil diameter was negatively correlated with the idiosyncratic model-derived belief strength (i.e., inverse uncertainty derived from subject-specific model fits) on the current trial about the source location, such that baseline pupil diameter was smaller when beliefs were more certain (Fig. 3a). After the tone was heard, the evoked pupil diameter was closely associated with surprise, with more surprising stimuli evoking larger pupil dilations with a time course similar to the switch modulations (Fig. 3b). Evoked change also reflected information about the prediction belief strength before having heard the tone, such that pupils dilated slightly less after stimulus onset if beliefs were very strong, regardless of surprise. Similar effects were evident in the subjects' behavioral response times (RTs), which tended to be shorter on trials when beliefs were more certain and longer on trials following the occurrence of a surprising stimulus (Fig. 3d,e). These RT effects, which also reflected a modest trial-to-trial correlation between RT and pupil size (median [interquartile range] Spearman's *rho* computed per subject= 0.11 [0.06–0.18] for baseline and 0.12 [0.07–0.18] for evoked pupil, sign-rank test for H_0: median=0, p=3.7x10^{-7} and 6.4x10^{-10},

respectively), testify to the behavioral relevance of these model-derived quantities but did not alone account for their relationships with pupil size: baseline and evoked pupil changes showed the same strong modulations by belief strength and surprise, respectively, even when accounting for trial-by-trial RTs (Fig. S2). Thus, on each trial, how new evidence was interpreted relative to expectations was reflected in the transient evoked pupil response, and how this information was used to update behaviorally relevant beliefs was reflected in baseline pupil diameter that was sustained across trials.

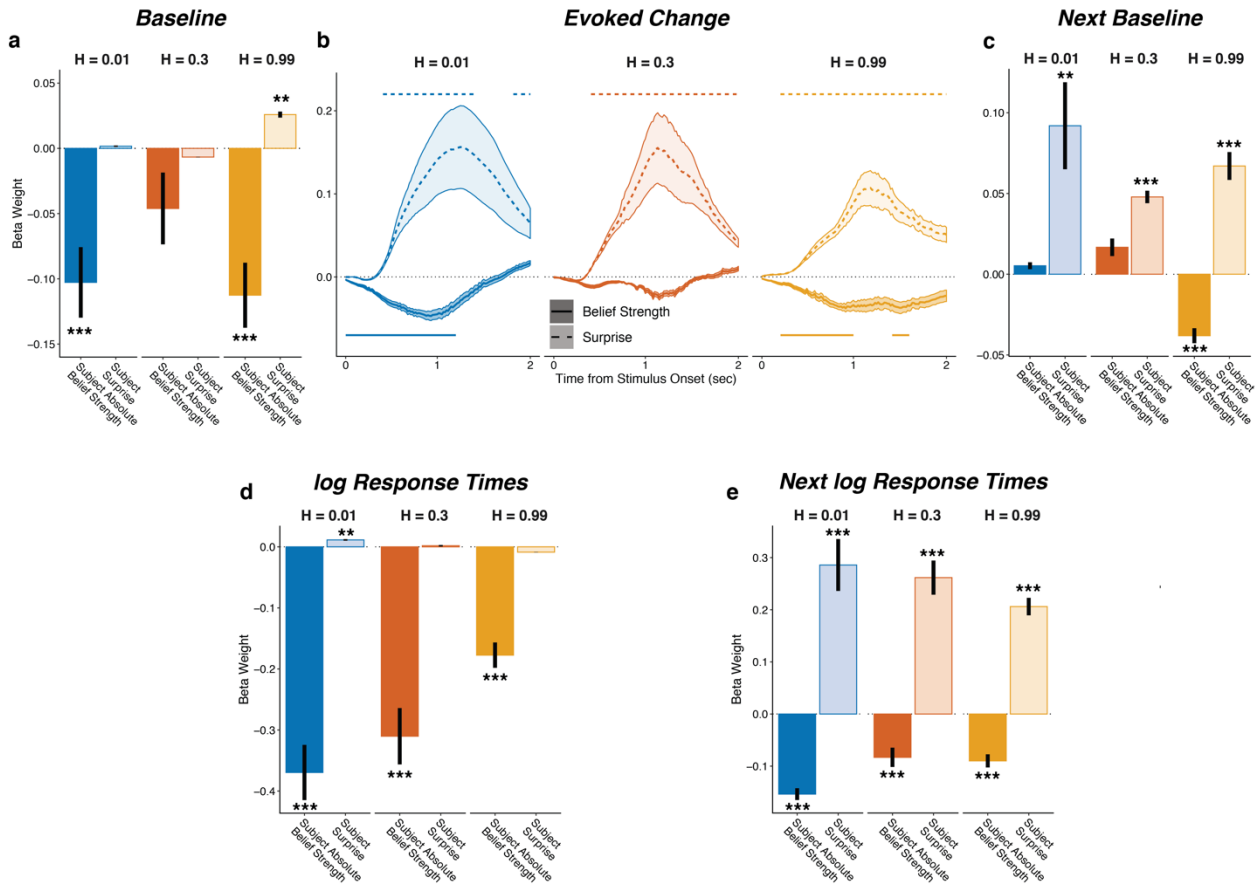

Figure 3. Pupils encode belief strength and surprise. Linear mixed effects beta weights estimating the influence of subject absolute belief strength and surprise on: a) baseline pupil diameter, b) evoked change, c) next baseline pupil diameter, d) log response times (i.e., the time taken to make a prediction before hearing the tone), and e) log response times on the next trial. Both baseline pupil measures accounted for the previous evoked pupil change. The evoked change and next pupil baseline measures accounted for the previous baseline pupil and the evoked pupil change and baseline pupil on the current trial (which accounts for the smaller modulation by belief strength in the next baseline, shown in c, compared to the current baseline, shown in a). Note that surprise depends on the current belief strength that was present before stimulus onset, which likely explains the small modulations by surprise in the high-hazard baseline pupil and low-hazard prediction RT. Bars and errorbars in a, c, d, and e or lines and shading in b correspond to mean±SEM across subjects. Straight lines above and below beta values in panel b indicate significant timepoints (FWE corrected $p<0.05$). In bar plots, **: $p<0.01$, ***: $p<0.001$.

Mental model complexity accounts for individual differences in belief-updating strategies and certain pupil dynamics

The fits of the Bayesian learning model captured individual variability in the subjects' belief-updating processes in terms of differences in the width of the prior distribution over hazard rate, which governs adaptability by establishing the range of hazard-rate "hypotheses" that can be inferred from the data (Fig. S1; Glaze et al., 2018). To assess more directly the information used by individual subjects to drive the idiosyncratic adaptability of their behavior and pupil dynamics, and to ensure that our conclusions are not overly dependent on the assumptions of the Bayesian learning model, we also analyzed behavior using a more general and model-agnostic information-bottleneck analysis (Fig. 4a; Palmer, Marre, Berry, & Bialek, 2013; Tishby, Pereira, & Bialek, 2000). We previously used this method to demonstrate that a subject's mental-model complexity influences adaptive decision-making in a way that is consistent with a bias-variance trade-off (Glaze et al., 2018). This approach assumes that a mental model Z compresses information from the past X_{past} to predict future observations X_{future}. Consistent with definitions of model complexity, more complex models, such as those that explore larger hypothesis spaces, encode more information from the past to make predictions (Fig. 4b; Bialek, Nemenman, & Tishby, 2001; Myung, Balasubramanian, & Pitt, 2000). Accordingly, complex models adapt flexibly to real environmental changes but risk overfitting spurious observations. In our task, each subject's model complexity corresponded to the amount of past information I_{past} encoded in Z and was measured by the mutual information between past observations and a current mental model, $I_{past} \equiv I(X_{past}; Z)$ (Chechik, Globerson, Tishby, & Weiss, 2005; Tishby et al., 2000). Unlike other methods that measure complexity based on a model's functional form (e.g., Bayesian model selection methods; Akaike, 1974; Myung et al., 2000; Schwartz, 1978), this method provides a principled and data-driven measure of mental-model complexity that requires no specific assumptions about the form of the model the subject is using to perform the task (Bialek et al., 2001; Chechik et al., 2005; Tishby et al., 2000).

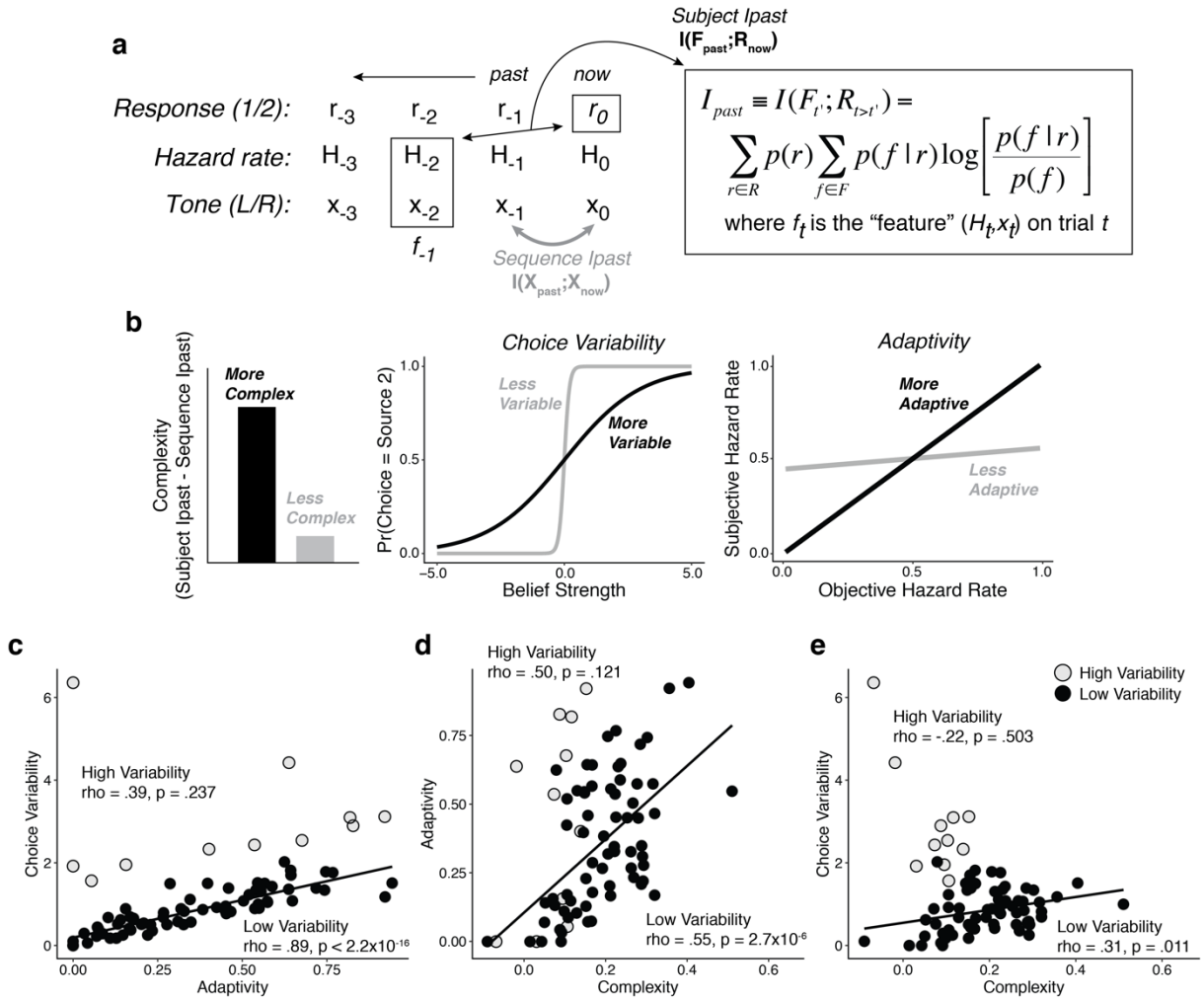

Figure 4. Subjects differ by mental-model complexity. a) The information-bottleneck measure of mental-model complexity, computed as the mutual information between past task features (the combination of the binary stimulus location, *x,* and hazard rate, *H,* from previous trials) on the current binary response (prediction), *r*. b) Schematic of the effects of high versus low complexity on choice variability (the slope of a psychometric function describing the probability of a particular choice as a function of belief strength) and adaptivity (the relationship between subjective and objective hazard rate)[19]. The bottom panels show for individual subjects (points): c) choice variability versus adaptivity, d) adaptivity versus complexity, and e) choice variability versus complexity. Lines are linear fits for cases for which H_0: Spearman correlation coefficient (*rho*, indicated for each data set)=0, $p<0.05$. Grey/black circles indicate data from high/low variability subjects, determined in terms of the offset from the regression line in c.

Consistent with our previous findings, subjects with wider fit priors over hazard rate used more complex mental models (Spearman correlation comparing log-prior-width with complexity: *rho*=-0.26, *p*=0.021, where lower log-prior-width values indicate wider priors). This increase in complexity, which implies that choice patterns were more strongly dependent on the hazard-rate condition, corresponded to increased adaptivity to real changes in hazard rate (i.e., low bias; Fig. 4d) but at the cost of more variable predictions (i.e., high variance; Fig. 4d). This relationship was particularly strong for a large subset of "low-variability" subjects whose behavior was more consistent with normative learning principles guided by priors of varying width than "high-variability" subjects who appeared to have additional sources of variability in their learning and/or decision processes (Fig. S3).

These individual differences in mental-model complexity, which like the fits of the Bayesian learning model were computed using the full, across-hazard sequence of trials for each subject, also reflected within-hazard differences in performance. In particular, subjects using higher overall mental-model complexity tended to use more flexible information processing over multiple trials and had higher overall performance within each hazard-rate condition (Fig. 5). For low hazard-rate trials, subjects using more complex mental models exploited the expected stability of the sound-source location by integrating information more strongly from past observations (Fig. 5a,b). As a result, their predictions better matched the objective hazard rate (assessed by fitting the normative choice model to behavior with subjective hazard rate as a free parameter in hazard-specific blocks; Glaze et al., 2015; Fig. 5c) and had higher accuracy (Fig. 5d). For intermediate hazard-rate trials, subjects using more complex mental models also integrated information more strongly across trials (Fig. 5a,b), thus treating the environment as slightly more stable than it was (Fig. 5c) but nonetheless resulting in modest accuracy gains (Fig. 5d). For high hazard-rate trials, subjects using more complex mental models tended to take advantage of the expected trial-by-trial switching of the sound-source location by weighing previous observations more strongly in accordance with this alternation to improve performance (Fig. 5). These effects all tended to occur for the low-variability subjects but not the high-variability subjects, further supporting the idea that our measure of mental-model complexity could both distinguish efficient (low-variability) from inefficient (high-variability) strategies and identify systematic differences in how efficient strategies used recent information to govern predictions.

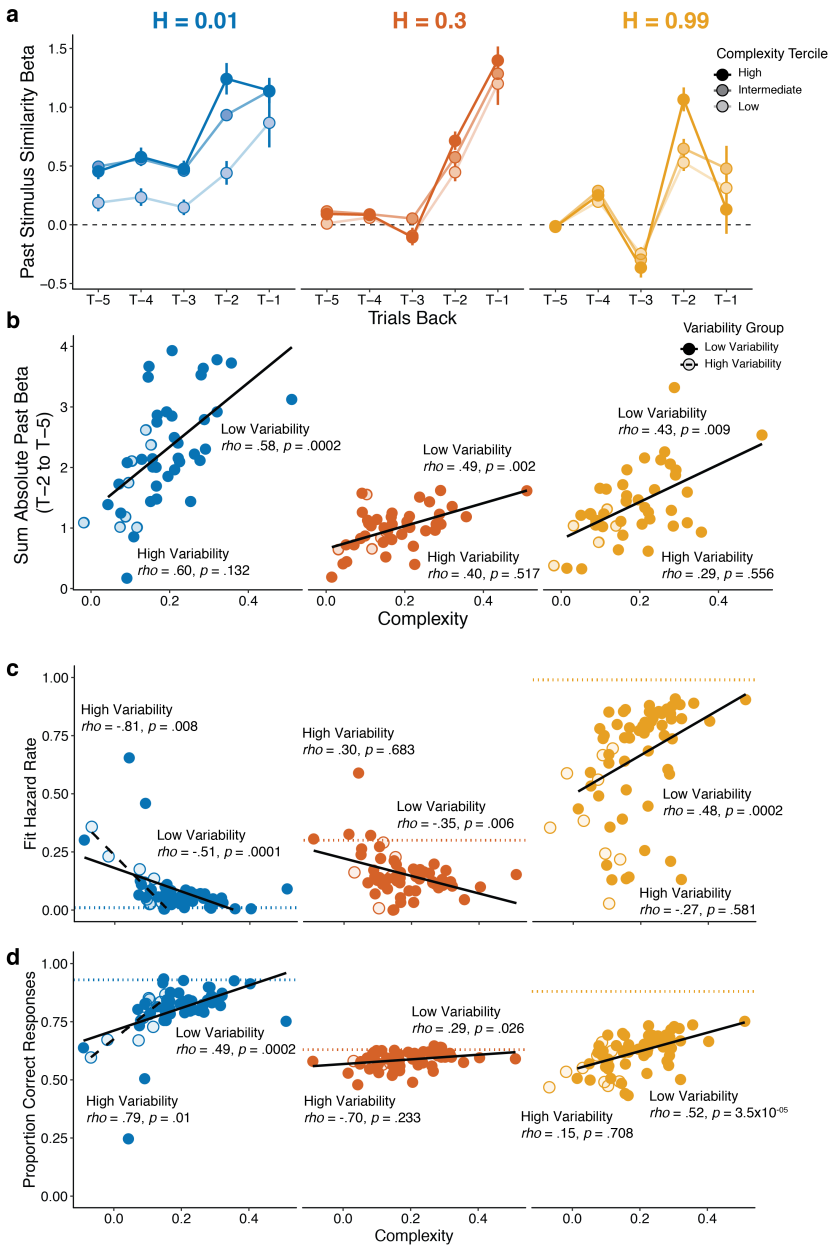

Figure 5. Individual differences in behavior. a) Average logistic regression beta weights measuring the influence of the similarity of stimuli up to five trials back on the current prediction per hazard rate for terciles of subjects with low, intermediate, and high complexity values. Positive/negative beta weights indicate a tendency to predict the same/opposite side as the stimulus location from the given trial. Points and error bars are mean±SEM across subjects. b–d) Complexity (abscissa) compared to: b) the sum of the past similarity beta weights 2–5 trials back, c) subjective hazard rate determined from fits of the Bayesian learning model (colored dotted lines correspond to objective hazard-rate values), and d) proportion correct responses (colored dotted lines correspond to mean proportion correct responses from the Bayesian learning model with a wide prior over hazard rate). Light/dark circles indicate data from high/low variability subjects. Lines are linear fits for cases for which H_0: Spearman correlation coefficient (*rho*, indicated for each data set)=0, $p<0.05$.

These analyses also suggest that the subjects had the most difficulty performing the task in the high hazard-rate condition, for several reasons. First, subjective hazard rates were most variable and most strongly mis-matched to objective values (dotted lines in Fig. 5c) in the high hazard-rate condition (*F*-test comparing the variance of fits for each hazard rate condition: low versus high, $p=2.2\text{x}10^{-10}$; intermediate versus high, $p=1.0\text{x}10^{-13}$; low versus intermediate, $p=0.179$). Second, performance accuracy was furthest from the ceiling provided by a fully adaptive Bayesian learning model in the high hazard-rate condition (dotted lines in Fig. 5d indicate the performance of the wide-prior model depicted in Fig. S1a). Third, RTs were slowest in the high-hazard condition (median RT for all trials and subjects = 465, 552, and 694 ms for the low, intermediate, and high hazard-rate condition, respectively; paired sign-rank test for H_0: median difference=0, $p<0.006$ for each paired comparison). Thus, despite the fact that the high hazard-rate condition provided the same predictive information as the low-hazard rate condition and more predictive information than the intermediate hazard-rate condition, subjects tended to make choices that were, on average, more consistent with lower hazard rates (Fig. S1b), with only the best-performing subjects developing appropriate expectations for both highly stable and highly unstable environments.

These findings, particularly the much wider range of subjective hazard rates in the high versus low and intermediate hazard-rate conditions (Fig. 5c), imply larger between-subject differences in the strategies used to treat stimulus switch versus non-switch trials in the high hazard-rate condition. Such across-subject differences in high-hazard belief updating should, in turn, reflect systematic differences in the pupil responses that encode the updating process in that condition. To test this idea directly, we assessed the relationship between mental-model complexity and the extent to which both belief updating and pupil responses differed between stimulus switch and non-switch trials for each hazard condition.

Using belief strength values derived from each subject's model fits, we computed the trial-by-trial belief that a switch had occurred and compared how these switch expectations differed across hazard conditions and trial types (stimulus switch versus no-switch trials). Both belief updating and pupil responses reflected stimulus switch versus non-switch differences as a function of mental-model complexity for the high hazard-rate condition alone (Fig. 6). Consistent with the other behavioral results, these relationships were evident only for low-variability subjects, for whom increasing mental-model complexity corresponded to an increasing understanding of the

statistical structure of the task: as complexity increased, stimulus switch trials became (appropriately) less surprising, which corresponded to smaller evoked pupil responses, and further reinforced the belief (i.e., reduced uncertainty) that the source had switched, which corresponded to smaller baseline pupil diameter (Fig. 6c-d). These effects were not evident in the low and intermediate hazard-rate conditions, for which differences in mental-model complexity involved more subtle belief adjustments to patterns of stimuli than simply better differentiating changes in beliefs following single stimulus switch versus non-switch trials. The high-hazard relationships to pupil diameter also did not result simply from RT differences across subjects (partial Spearman correlations relating both complexity and median RT to pupil: *rho*=-0.43, *p*=0.006, for complexity versus baseline pupil; -0.36, *p*=0.023, for complexity versus evoked pupil; 0.11, *p*=0.532, for RT versus baseline pupil; -0.11, *p*=0.504, for RT versus evoked pupil). Finally, individual differences in mental-model complexity were not related to average values (i.e., as opposed to stimulus switch versus non-switch differences as in Fig. 6) of baseline or evoked pupil for any hazard-rate condition (Fig. 7). Together, these findings support the idea that cognitive modulations of pupil diameter are not simply static markers of overall effort but rather dynamic signals representing if and how that effort is used to update beliefs about the world given new, context-dependent information.

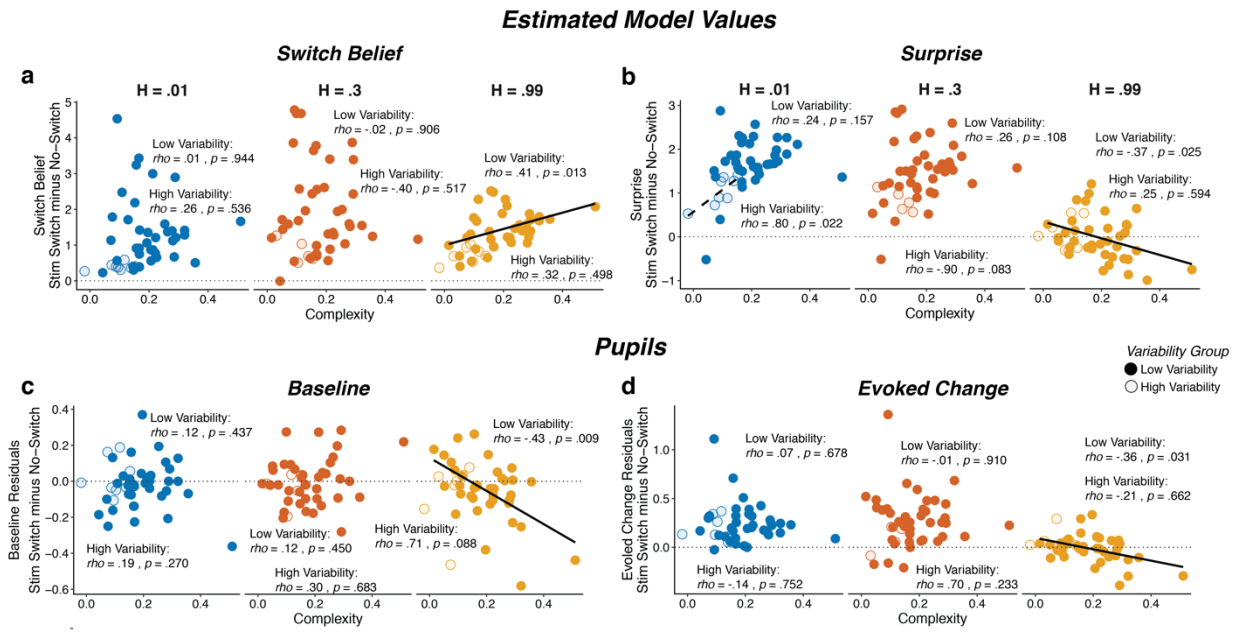

Figure 6. Individual differences in mental-model complexity represented in pupil contrasts in the high-hazard condition. All panels show data from individual subjects (light/dark circles indicate data from high/low variability subjects) separated by hazard-rate condition (colors, as indicated) and plotted as a function of the subject's mental model complexity (abscissa). The top panels show how belief that a switch occurred (a) or surprise (b), assessed from fits of the Bayesian learning model to behavioral data from each subject, differed just following a stimulus switch versus non-switch trial. Positive values imply that a stimulus switch was more likely to strengthen the belief that the stimulus switched (a) and was surprising (b). The bottom panels show how the pupils responded to stimulus switch versus non-switch trials, either in the following baseline epoch (c) or in terms of the evoked response to the tone (d). Pupil measurements were computed as residuals from a linear model that accounted for nuisance variables including the baseline pupil diameter and evoked change in pupil diameter from the previous trial (Eq. 13). Lines are linear fits for cases for which H_0: Spearman correlation coefficient (*rho*, indicated for each data set)=0, $p<0.05$.

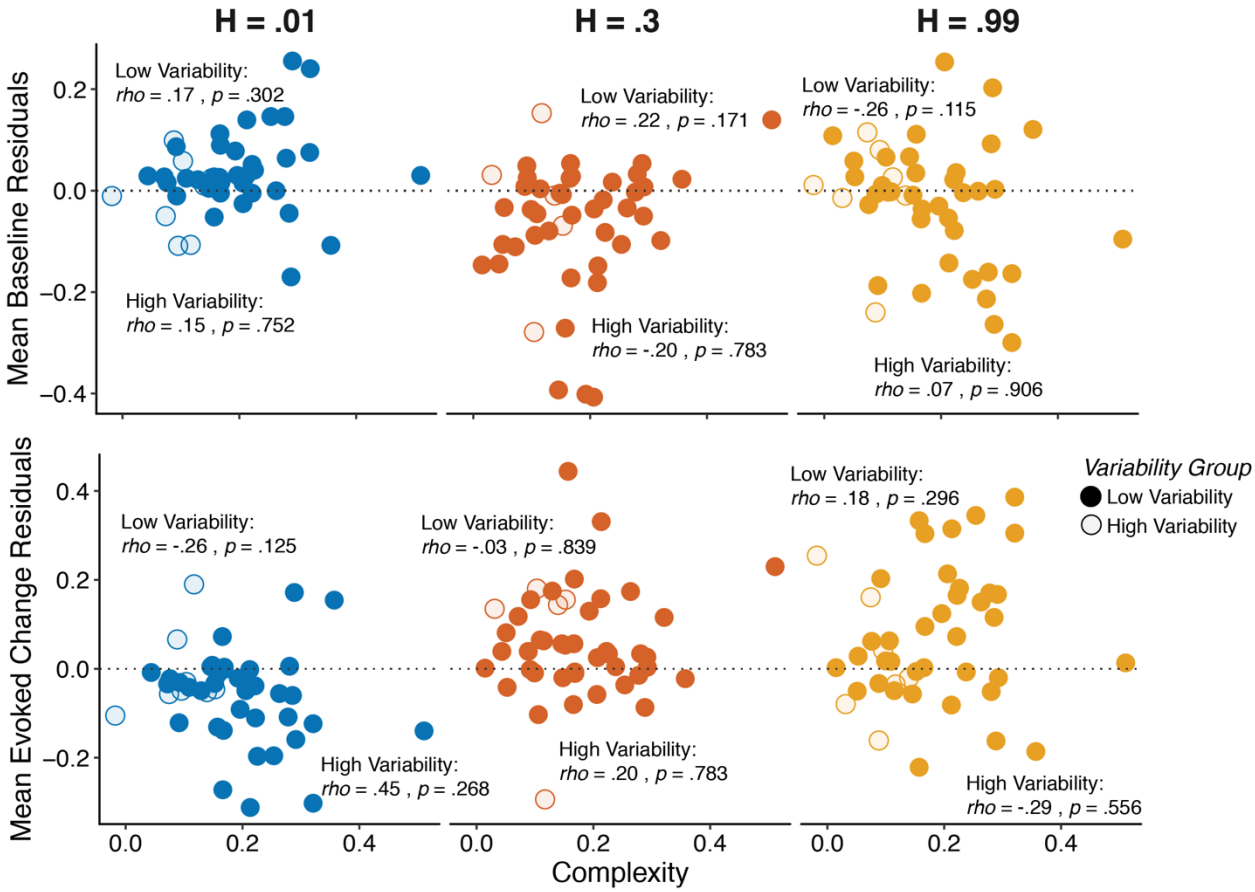

Figure 7. Individual differences in mental-model complexity were not represented in overall pupil responses. All panels show data from individual subjects (light/dark circles indicate data from high/low variability subjects) separated by hazard-rate condition (colors, as indicated) and plotted as a function of the subject's mental model complexity (abscissa). The ordinate corresponds to baseline (top) or evoked (bottom) pupil diameter across all trials. Both pupil measurements were computed as residuals from a linear model that accounted for nuisance variables including the baseline pupil diameter and evoked change in pupil diameter from the previous trial (Eq. 13). All correlation *p*-values>0.05.

Discussion

We examined the link between dynamics in pupil diameter and belief changes during an on-line adaptive learning task and report three primary findings. First, baseline and evoked changes in pupil diameter reflect context-dependent violations of learned expectations and not just bottom-up stimulus changes. Whereas previous research found these signals in relatively stable (e.g., low-hazard) task environments, we show that they are also present in environments that change frequently, but predictably. Second, these pupil dynamics reflect learned inferences about the latent task structure (i.e., the hazard rate). This learning process is consistent with a model that uses surprise, which is reflected in evoked pupil diameter, to update a belief about the current state of the task, the strength of which is reflected in baseline pupil diameter. Third, these pupil dynamics also reflect individual differences in the ability to generate appropriate, context-dependent top-down expectations, which we measured as mental-model complexity. These complexity differences can affect both baseline and evoked pupil responses, especially for task conditions in which both belief strength and surprise, respectively, are affected systematically by complexity.

These results help to clarify the nature of cognitive modulations of pupil diameter, including baseline values that have been associated with "mental effort" (Beatty, 1982; Mathôt, 2018). According to our results, these modulations do not simply reflect task difficulty in terms of: 1) objective difficulty, which was equivalent in the low and high hazard-rate conditions that had substantially different pupil responses; 2) overall performance, which was similar in the intermediate and high hazard-rate conditions that had substantially different pupil responses; or 3) RTs, which were only weakly related to pupil diameter on a trial-by-trial basis. Moreover, these modulations are not simply overall offsets in baseline or evoked pupil diameter during periods of high cognitive load, because these context-independent pupil metrics were largely unaffected in our study. Instead, our data suggest that cognitively driven pupil responses are dynamic signals related closely to idiosyncratic belief updating, such that the complexity, and perhaps effort, associated with making effective use of context to process incoming sensory information is associated with corresponding context-dependent modulations of pupil diameter.

Our findings provide insights into the neural mechanisms that underlie this kind of adaptive and dynamic belief-updating process. Increases in pupil diameter have been tied to higher levels of activation of the brainstem nucleus locus coeruleus (LC), which is the primary source of

norepinephrine (NE) to the brain (de Gee et al., 2017; Joshi, Li, Kalwani, & Gold, 2016; Liu, Rodenkirch, Moskowitz, Schriver, & Wang, 2017; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014). Adaptive Gain Theory proposes that tonic release of NE occurs during periods of active exploration (Aston-Jones & Cohen, 2005; Aston-Jones, Rajkowski, & Cohen, 1999). This idea is consistent with our finding of increased baseline pupil under conditions of low belief (high uncertainty). Specifically, although physical exploration is absent in our task, there may be a form of "mental exploration" that occurs when beliefs are weak and alternative explanations for the observed data are sought in terms of the hypothesis space of possible hazard rates (e.g., the prior distribution over hazard rates in our Bayesian learning model; Collins & Koechlin, 2012; Glaze et al., 2018; Wilson, Nassar, & Gold, 2010). Our results suggest that this covert process, like active exploration in Adaptive Gain Theory, may involve increases in activation of the LC-NE system during periods of high uncertainty(Yu & Dayan, 2005).

Phasic, event-driven NE release, possibly corresponding to evoked pupil changes in response to surprise as described in our and other studies (Joshi et al., 2016; Nassar et al., 2012; O'Reilly et al., 2013; Preuschoff et al., 2011) may play a complementary role in adaptive inference. Instead of promoting exploration, phasic NE release is thought to result in widespread network "resets" that interrupt and initiate the reorganization of neural networks engaged during a task, possibly reducing previously learned biases and enhancing the relevance of new observations (Bouret & Sara, 2005; de Gee et al., 2017; Sara, 2009). Our results suggest that this kind of network reset can occur for not just an abrupt, surprising change in sensory input, but also an abrupt, surprising lack of change in sensory input.

Sensitivity to such different kinds of violations of expectations would require the LC and other relevant brain systems to have access to a flexible, complexity-dependent combination of bottom-up and top-down information. One possible source of such information is the anterior cingulate cortex (ACC ;Aston-Jones & Cohen, 2005), which is reciprocally connected with the LC and has been proposed to guide executive control, allocating when and how executive resources are used during cognitively demanding tasks (Filipowicz et al., 2016; Shenhav, Botvinick, & Cohen, 2013; Shenhav, Cohen, & Botvinick, 2016; Shenhav et al., 2017). During adaptive decision-making, the ACC and the anterior insula are active when changes are detected and mental models are updated (Behrens et al., 2007; Donoso, Collins, & Koechlin, 2014; McGuire et al., 2014; O'Reilly et al., 2013; Sarafyazd & Jazayeri, 2019; Stöttinger et al., 2015). Moreover, these

kinds of signals have been linked directly with computations related to surprise and uncertainty (McGuire et al., 2014; Nassar, Mcguire, Ritz, & Kable, 2019). The ACC and anterior insula mediate activation of prefrontal cortical regions, such as the anterior and dorsolateral prefrontal cortex (Bastin et al., 2017; Domenech & Koechlin, 2015; Ham, Leff, de Boissezon, Joffe, & Sharp, 2013; Menon & Uddin, 2010; Stöttinger et al., 2015), that are involved in the prospective and counterfactual reasoning that are characteristic of the exploratory strategies humans use to update mental models (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Domenech & Koechlin, 2015; Donoso et al., 2014; Koechlin & Hyafil, 2007; Koechlin & Summerfield, 2007; Stöttinger et al., 2015; Zajkowski, Kossut, & Wilson, 2017). Our results suggest that increased activity in the LC, along with the network of brain regions with which it connects, could be more engaged in subjects with more complex mental models, guiding some of the idiosyncratic strategies observed between subjects.

Finally, our task highlighted a learning asymmetry between the low and the high hazard-rate conditions, with subjects learning better in predictably stable environments than from predictably unstable environments. This bias was evident in the average group prior over hazard rate, which was centered ~0.3 and had little mass around higher hazard values (Fig. S1b). This asymmetry was also mirrored in the way stimulus properties were reflected in evoked pupil changes. Pupil dilations to rare "oddball" events have been reported extensively[5] but primarily for conditions similar to our low hazard-rate trials, in which stimulus switches occur infrequently [8,59]. In our study, evoked changes associated with stimulus switches were similar on low and intermediate hazard-rate trials, but not for rare stimulus repeats on high hazard-rate trials, even though these non-switch events should be as computationally surprising as switch events in the low hazard-rate condition. Phasic arousal has been linked with decreases in choices biases, serving as a type of interrupt signal (Aston-Jones et al., 1999; de Gee et al., 2017). We propose that low-level arousal signals related to phasic pupil dilation may interact with expectation driven-signals to strengthen the system's ability to detect and adapt to changes. When these low-level signals are absent, as they were in response to repeat trials in the high hazard-rate condition, learning may rely more on top-down expectation driven signals and adapt less effectively.

In summary, the results of this study provide new insights into the cognitive signals that drive pupil diameter, which include both bottom-up and top-down processing that is used in an idiosyncratic way to update and maintain appropriate beliefs about an environment. These results

can help guide future research to examine how dynamic modulation of physiological arousal influences the information processing trade-offs humans use when learning in dynamic and uncertain environments.

Methods

Auditory prediction task

Seventy-eight University of Pennsylvania undergraduates (49 male, 29 female) performed an auditory prediction task in which they attempted to predict which of two auditory sources were generating tones on each trial. Subjects were presented with a blue screen and white fixation circle. On each trial, subjects made a prediction via left and right buttons on a game pad as to which source (left or right) they thought would generate the next tone. Three seconds after each prediction, a 300 ms auditory tone (196 Hz) was played in either the left or right ear via headphones. The next trial then started with no inter-trial interval; thus, RTs were measured from the onset of the tone on the previous trial to the onset of the button press on the current trial. Subjects were encouraged to maintain fixation throughout the task but were allowed self-timed breaks by breaking fixation after making a prediction. Visual and auditory stimuli were generated using SnowDots stimulus presentation software in MATLAB (https://github.com/TheGoldLab/Lab-Matlab-Control).

Subjects were informed that the left source had an 80/20 probability of generating left/right tones, respectively, and that the right source had an 80/20 probability of generating right/left tones, respectively. Subjects were also informed that one source was generating tones on each trial, and that the sources would sometimes switch. Subjects were also made aware that the rate of these switches, or 'hazard rate', would also switch, such that on some trials, the sources switched either infrequently (H=0.01; low hazard rate), at an intermediate rate (H=0.3; intermediate hazard rate), or very frequently (H=0.99; high hazard rate). Subjects performed 1–3 sessions of 1000 trials each. Each session occurred on separate days, with two of the three potential hazard rates switching off in each session (Fig. 1b).

Session order was counterbalanced between subjects. However, the sequence of tones presented in each session was identical, such that random patterns appearing within the sequence of tones were held constant between subjects. Subjects received $10 USD per hour of participation

for participation with an additional performance bonus for accumulated correct predictions (total per session payout was $10–20 USD).

The study protocol was reviewed and approved by the University of Pennsylvania's Institutional Review Board. All subjects provided informed consent prior to participating in the study.

Measuring pupil diameter

Of the 78 total subjects, 55 had their pupils recorded while performing the task. Pupils were recorded using either a Tobii T60-XL sampling both eyes at 60 Hz, or an SR Research Eyelink 1000+ sampling the right eye at 1000 Hz. Signals from the Eyelink were subsequently down sampled to 60 Hz to make them comparable to signals from the Tobii. Baseline pupil diameter was measured as the mean pupil diameter 0–100 ms prior to tone onset. Evoked changes were measures as baseline-subtracted pupil diameter during a 2 s period after stimulus onset. Trial-by-trial evoked changes in Figs. 6 and 7 were computed from the baseline-subtracted peak pupil diameter after stimulus onset.

Pupil processing

Pupil samples corresponding to blinks and abnormally fast changes in pupil diameter (e.g., around blink events) were removed, as were any additionally samples in which the eye position fell 10% of the screen diameter away from the fixation point. Eye position was additionally regressed from all data while preserving the intercept to estimate baseline pupil diameter. Missing values were then linearly interpolated and filtered using a first order Butterworth filter with a 4 Hz cutoff (Browning, Behrens, Jocham, O'Reilly, & Bishop, 2015; Krishnamurthy et al., 2017; Nassar et al., 2012). Trials with >50% missing/interpolated samples were omitted from further analysis (~9% of trials were omitted; Browning et al., 2015).

Complexity measure

We used a data-driven approach to measure mental-model complexity based on the principles of predictive information and the information bottleneck (Bialek et al., 2001; Glaze et al., 2018; Tishby et al., 2000). This information-bottleneck method seeks the smallest compression Z of past observations that retain maximum predictability about the future. This compression

effectively represents the model encoding this past information, and the size of the compression corresponds to the model's complexity (Chechik et al., 2005; Tishby et al., 2000). The size of the compression of past information I_{past} is measured by the mutual information between the Z and past observations X_{past}, $I_{past} \equiv I(X_{past}; Z)$.

Mental-model complexity was measured as the extent to which a subject's responses R (a prediction of the left or right source) on the current trial t encoded past stimulus features F from past trials $t' < t$:

$$I_{past} \equiv I(F_{t'<t}; R_t) = \sum_{r \in R} \sum_{f \in F} \Pr(r, f) \log_2 \left[\frac{\Pr(r, f)}{\Pr(r) \Pr(f)} \right] \tag{1}$$

The features F consisted of all six possible combinations of tone side (left or right) and hazard rate (0.01, 0.3, or 0.99) on each trial. Information-theoretic quantities such as mutual information can be biased due to small sampling errors. Although this influence was mitigated in part by the fact that the sequence of tones was held constant in each session for each subject, not all subjects participated in all sessions. We therefore corrected for these biases by subtracting the mutual information contained within the sequence of tones themselves, $I(tones_{t'<t}; tones_t)$, from each subject's I_{past} (Glaze et al., 2018). The remaining information corresponds to the amount of information subjects encode about the task hazard rate, beyond that provided by the stimuli.

Adaptivity model

We fit a descriptive adaptivity model that has previously been used to describe how effectively subjects incorporate changes in a task's hazard rate (Glaze et al., 2018, 2015). For this task, the subject's goal was to predict which of two sources will generate a tone on the current trial. The adaptivity model characterizes the belief L on trial t as the posterior log-odds that the next tone will be generated by the first or second source:

$$L_t = \psi_t(L_{t-1}, H_t) + LLR_t \tag{2}$$

where ψ_t indicates the log-prior-odds that either source will generate the next tone and LLR_t indicates the log evidence that the tone was generated by the first or second source. Given that there are only two states, we assigned positive log-prior-odds to the probability that a tone would be generated by the right source. The log-prior-odds ψ were computed using the log-posterior-odds from the previous trial L_{t-1} and incorporating the current hazard rate H_t:

$$\psi_t(L_{t-1}, H_n) = L_{t-1} + \log\left[\frac{1-H_t}{H_t} + \exp(-L_{t-1})\right] - \log\left[\frac{1-H_t}{H_t} + \exp(L_{t-1})\right] \quad (3)$$

The posterior belief for a trial t was then calculated as the log-prior-odds plus the evidence provided on each trial, calculated as the log-likelihood ratio that the current tone (x_t) was generated by the left or right source (z_1 and z_2 respectively)

$$LLR_t = \log\left[\frac{\Pr(x_t|z_1)}{\Pr(x_t|z_2)}\right] \quad (4)$$

This model characterizes how effectively subjects changed their subjective estimates of the current hazard rate throughout the task by calculating subjective estimates of the current log hazard-rate odds $J_{t,subjective}$ as a linear function of the current, objective log-hazard rate odds $J_{t,objective}$

$$J_{t,subjective} = J_{default} + m_H \times J_{t,objective} \quad (5)$$

where J indicates the log-odds of a hazard rate H (i.e., $J = \log(H/(1-H))$). Here, $J_{default}$ indicates the intercept and m_H indicates the slope of the regression, which describes the subject's adaptivity to changes in the task hazard rate.

The model was fit to choice behavior by writing the probability of choosing the right source as a logistic function of the log-prior-odds ψ_t as the decision variable, given that the subject's responses were made before hearing a tone on each trial, and a free parameter v capturing noise in the decision variable (i.e., choice variability):

$$\Pr(\hat{c}_t) = 1/(1 + \exp(-\psi_t/v)) \tag{6}$$

Fitting was performed using gradient decent with multiple starting points to find values of the three free parameters, ($J_{default}$, m_H, and v) that minimize the cross entropy:

$$e = -\sum_t (1 - c_t)\log\big(1 - \Pr(\hat{c}_t)\big) + c_t \log\big(\Pr(\hat{c}_t)\big) \tag{7}$$

Although a lapse term was previously included in this model (Glaze et al., 2018, 2015), adding this term did not improve our model fits and thus was not included here (BICs were lower for a model with choice variability and no lapse compared to a model with choice variability and lapse in 63 of 78 subjects; paired signed-rank test comparing BICs: p=0.0004). Subjective estimates of each separate hazard rate (Fig. 5c) were computed by fitting the same model in Eq. 2 separately for each hazard-rate condition, keeping H as a free parameter in each case.

High-Variability Subjects

Although choice variability was generally within the range observed in our previous studies (maximum choice variability in previous studies=1.02 (Glaze et al., 2018, 2015; compare to values in Fig. 4c) a number of subjects had far higher choice variability. To identify these subjects, we applied the generalized extreme studentized deviate many-outliers test (Rosner, 1983) (implemented using the PMCMRplus package in R) on the residuals of the linear relationship between adaptivity and choice variability (Fig. 4c). This procedure identified 11 subjects with outlying variability scores, which are identified throughout as "high-variability" subjects. All remaining subjects are identified as "low-variability" subjects).

Bayesian learning model

We used the same hierarchical Bayesian on-line inference algorithm used previously as a normative model for the current task (Glaze et al., 2018). This model predicts which of the two sources *z* is most likely to generate a tone on the next trial. This model assumes that the likelihood that the sources switch is governed by a hazard rate (*H*), and that there is a fixed probability *K* that *H* changes from trial to trial.

For each subject, H is assumed to be generated from a beta distribution $\Pr_0(H) = Beta(\alpha, \beta)$ with mean μ and precision ϕ parameters, such that $\alpha = \mu\phi$ and $\beta = (1-\mu)\phi$. The goal was to actively infer which of the two possible sources *z* (right or left) would generate the next tone *x*. Tones were $Bernoulli(p)$ distributed, with *p*=0.8. To infer the next tone, we sought a distribution over source-space for a given trial *t*, using all of the previously heard tones thus far $\Pr(z_t|x_{1:t-1})$:

$$\Pr(H_t, z_t|x_{1:t-1}, \mu, \phi, K, p) = \frac{\Pr(H_t, z_t, x_{1:t-1}|\mu, \phi, K, p)}{\int_0^1 dH_t \sum_{z_{t-1}} \Pr(H_t, z_t, x_{1:t-1}|\mu, \phi, K, p)} \tag{8}$$

The numerator can be re-written recursively as:

$$\Pr(H_t, z_t, x_{1:t-1}|, \mu, \phi, K, p) = \int_0^1 dH_{t-1} \Pr(H_t|H_{t-1}, K) \sum_{z_{t-1}} \Pr(z_t|z_{t-1}, H_t) \Pr(H_{t-1}, z_{t-1}, x_{1:t-1}|\mu, \phi, K, p) \tag{9}$$

Where the transition probabilities in hazard-rate space are defined as

(10)

$$\Pr(H_t|H_{t-1},K) = \begin{cases} 1-K+K\Pr_0(H_t) & H_t = H_{t-1} \\ K\Pr_0(H_t) & H_t \neq H_{t-1} \end{cases}$$

and the transition probabilities in source space as

$$\Pr(z_t|z_{t-1},H_t) = \begin{cases} (1-H_t) & z_t = z_{t-1} \\ H_t & z_t \neq z_{t-1} \end{cases} \tag{11}$$

The normative model's belief L_t about the location of the source on trial *t* after hearing a tone was computed as the log-posterior-odds of each state

$$L_t = LLR_t + \psi_t$$

$$LLR_t = \log\left[\frac{\Pr(x_t|z_1)}{\Pr\,(x_t|z_2)}\right] \tag{12}$$

$$\psi_t = \log\left[\frac{\int_0^1 dH_t \int_0^1 H_{t-1}\Pr(H_t|H_{t-1},K)\sum_{z_{t-1}}\Pr(z_{1t}|z_{1:t-1},H_n)\Pr(H_{t-1},z_{t-1},x_{1:t-1}|\mu,\phi,K,p)}{\int_0^1 dH_t \int_0^1 H_{t-1}\Pr(H_t|H_{t-1},K)\sum_{z_{t-1}}\Pr(z_{2t}|z_{1:t-1},H_n)\Pr(H_{t-1},z_{t-1},x_{1:t-1}|\mu,\phi,K,p)}\right]$$

This model was fit with four free parameters: μ, ϕ, K governing the learning process and an additional choice variability parameter, v, identical to the one in Eq. 4. The model was fit by minimizing the same cross entropy function in Eq. 5. As with the descriptive adaptivity model, a lapse parameter was omitted, because adding this term did not improve our fits (BICs were lower for the choice variability only model for 65 of 78 subjects; paired signed-rank test comparing BICs: p=4.5x10^{-5}).

For the regression analyses, absolute belief strength was computed as $|\psi_t|$ and surprise was computed after applying eq. 6 as $-\log\,[\Pr(\hat{c})]$ or $-\log\,[1-\Pr(\hat{c})]$ for tone played in the right or left ear respectively. Belief in source switch after hearing a tone (Fig.6a) was computed as

$\psi_{t-1} - \psi_t$ when $\psi_{t-1} > 0$, and $\psi_t - \psi_{t-1}$ when $\psi_{t-1} < 0$, such that positive/negative values indicate belief in a source switch/no-source switch respectively.

Individual differences analyses

Individual differences in pupil responses were calculated using mixed-effects models implemented with the "lme4" package (Bates, Mächler, Bolker, & Walker, 2015) in the R statistical language (R Core Team, 2019). These models included fixed variables of interest (i.e., stimulus and prediction switch variables, log RTs, and model parameters), along with known nuisance variables $\beta_{nuisance}$ (Krishnamurthy et al., 2017) and random slopes and intercepts u:

$$Pupil = \beta_0 + \beta_{1-n} + \beta_{nuisance} + u_0 + u_{1-n} + u_{nuisance} \quad (13)$$

When computing the influences of factors on baseline pupil diameter, the baseline and baseline-subtracted evoked change were included as nuisance parameters, with the exception of analyses examining the influence of absolute belief strength (i.e., Fig. 3a and S2a), which only accounted for previous evoked changes. This was done to account for the fact that absolute belief strength is a cumulative variable, depending in part on the belief strength from the previous trial. When measuring evoked changes and next baseline, baseline pupil diameter from the current trial was also added as a nuisance variable (Krishnamurthy et al., 2017). Individual subject differences were obtained by examining the individual slopes for each variable of interest.

Significance testing for the linear mixed effects models was done using the "lmerTest" package (Kuznetsova, Brockhoff, & Christensen, 2017), which computes *p* values for *F* and *t* statistics using Satterthwaite's method for approximating degrees of freedom (Fai & Cornelius, 1996).

Acknowledgments

The authors would like to thank Kenan Saleh and Mohammed Kabir for their assistance with data collection. Funded by NSF-NCS 1533623, R01 EB026945, and NIMH F32 MH117924. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

C.M.G., J.I.G & J.W.K designed the experiment; A.L.S.F. performed all of the analyses and computational modeling; all four authors interpreted the results and drafted and/or revised the manuscript.

Competing interests

The authors declare to competing interests.

References

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Aston-Jones, G., & Cohen, J. D. (2005). An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry*, *46*(9), 1309–1320. https://doi.org/10.1016/S0006-3223(99)00140-7

Bastin, J., Deman, P., David, O., Gueguen, M., Benis, D., Minotti, L., … Jerbi, K. (2017). Direct Recordings from Human Anterior Insula Reveal its Leading Role within the Error-Monitoring Network. *Cerebral Cortex*, *27*, 1545–1557. https://doi.org/10.1093/cercor/bhv352

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bialek, W., Nemenman, I., & Tishby, N. (2001). Predictability , Complexity , and Learning. *Neural Computation*, *13*(11), 2409–2463.

Bouret, S., & Sara, S. J. (2005). Network reset : a simplified overarching theory of locus coeruleus noradrenaline function, *28*(11). https://doi.org/10.1016/j.tins.2005.09.002

Bremner, F. (2009). Pupil evaluation as a test for autonomic disorders. *Clinical Autonomic Research*, *19*(2), 88–101. https://doi.org/10.1007/s10286-009-0515-2

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, *18*(4), 590–596. https://doi.org/10.1038/nn.3961

Chechik, G., Globerson, a, Tishby, N., & Weiss, Y. (2005). Information bottleneck for Gaussian variables. *J Mach Learn Res*, *6*(1), 165–188.

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLoS Biology*, *10*(3). https://doi.org/10.1371/journal.pbio.1001293

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

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*, (e23232), 1–36. https://doi.org/10.7554/eLife.23232

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences*, *1*, 101–106. https://doi.org/10.1016/j.cobeha.2014.10.007

Donoso, M., Collins, A. G. E., & Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science*, *344*(6191), 1481–1486. https://doi.org/10.1126/science.1252254

Fai, A. H.-T., & Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation*, *54*(4), 363–378. https://doi.org/10.1080/00949659608811740

Faraji, M. (2018). Balancing New against Old Information : The Role, *83*, 34–83. https://doi.org/10.1162/NECO

Filipowicz, A., Anderson, B., & Danckert, J. (2016). Adapting to change: The role of the right

hemisphere in mental model building and updating. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(3), 201–218. https://doi.org/10.1037/cep0000078

Filipowicz, A., Valadao, D., Anderson, B., & Danckert, J. (2014). Measuring the influence of prior beliefs on probabilistic estimations. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*, 2198–2203. https://doi.org/10.13140/2.1.1323.8406

Filipowicz, A., Valadao, D., Anderson, B., & Danckert, J. (2018). Rejecting Outliers: Surprising Changes Do Not Always Improve Belief Updating. *Decision*, *5*(3), 165–176. https://doi.org/10.1037/dec0000073

Glaze, C. M., Filipowicz, A. L. S., Kable, J. W., Balasubramanian, V., & Gold, J. I. (2018). A bias-variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nature Human Behaviour*, *2*(3), 213–224. https://doi.org/10.1038/s41562-018-0297-4

Glaze, C. M., Kable, J., & Gold, J. (2015). Normative evidence accumulation in unpredictable environments. *ELife*, *4*, e08825. https://doi.org/10.7554/eLife.08825

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*(1), 10–16. https://doi.org/10.1016/S1364-6613(00)01567-9

Ham, T., Leff, A., de Boissezon, X., Joffe, A., & Sharp, D. J. (2013). Cognitive Control and the Salience Network: An Investigation of Error Processing and Effective Connectivity. *Journal of Neuroscience*, *33*(16), 7091–7098. https://doi.org/10.1523/JNEUROSCI.4692-12.2013

Johnson-Laird, P. N. (2004). The history of mental models. In K. Manktelow & M. C. Chung (Eds.), *Psychology of reasoning: Theoretical and historical perspectives* (pp. 179–212). Hove, Sussex: Psychology Press.

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

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science (New York, N.Y.)*, *318*(5850), 594–598. https://doi.org/10.1126/science.1142995

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*(6), 229–235. https://doi.org/10.1016/j.tics.2007.04.005

Krishnamurthy, K., Nassar, M. R., Sarode, S., & Gold, J. I. (2017). Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nature Human Behaviour*, *1*(6). https://doi.org/10.1038/s41562-017-0107

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Liao, H., Yoneya, M., Kidani, S., & Kashino, M. (2016). Human Pupillary Dilation Response to Deviant Auditory Stimuli : Effects of Stimulus Properties and Voluntary Attention. *Frontiers in Neuroscience*, *10*, 43. https://doi.org/10.3389/fnins.2016.00043

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*, *20*(13), 3099–3112. https://doi.org/10.1016/j.celrep.2017.08.094

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(47), 12539–12545. https://doi.org/10.1523/JNEUROSCI.2925-08.2008

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

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870–881. https://doi.org/10.1016/j.neuron.2014.10.013

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, *214*(5–6), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Muller, T. H., Mars, R. B., Behrens, T. E., & Reilly, J. X. O. (2019). Control of entropy in neural models of environmental state. *ELife*, *8*(e39404), 1–30.

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

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences*, *97*(21), 11170–11175. https://doi.org/10.1073/pnas.170283897

Nassar, M. R., Mcguire, J. T., Ritz, H., & Kable, J. W. (2019). Dissociable Forms of Uncertainty-Driven Representational Change Across the Human Brain, *39*(9), 1688–1698.

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

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

O'Reilly, J. X. (2013). Making predictions in a changing world-inference, uncertainty, and learning. *Frontiers in Neuroscience*, *7*(7 JUN), 1–10. https://doi.org/10.3389/fnins.2013.00105

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*, *110*(38), E3660–E3669. https://doi.org/10.1073/pnas.1305373110

Palmer, S. E., Marre, O., Berry, M. J., & Bialek, W. (2013). Predictive information in a sensory population. *Proceedings of the National Academy of Sciences*, *112*(22), 6908–6913. https://doi.org/10.1073/pnas.1506855112

Preuschoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, *5*(SEP), 1–12. https://doi.org/10.3389/fnins.2011.00115

Qiyang, J., Richer, F., Wagoner, B., & Beatty, J. (1985). The pupil and stimulus probability. *Psychophysiology*, *55*(5), 530–534.

Rosner, B. (1983). Percentage points for a generalized esd many-outlier procedure. *Technometrics*, *25*(2), 165–172. https://doi.org/10.1177/0892020607085626

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), 211–223. https://doi.org/10.1038/nrn2573

Sarafyazd, M., & Jazayeri, M. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. *Science*, *364*(6441), 1–9. https://doi.org/10.1126/science.aav8911

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Cohen, J. D., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, *19*(10), 1280–1285. https://doi.org/10.1038/nn.4382

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, *40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Stöttinger, E., Filipowicz, A., Valadao, D., Culham, J. C., Goodale, M. A., Anderson, B., & Danckert, J. (2015). A cortical network that marks the moment when conscious representations are updated. *Neuropsychologia*, *79*, 113–122. https://doi.org/10.1016/j.neuropsychologia.2015.10.037

Team, R. C. (2019). R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing,.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science (New York, N.Y.)*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Tishby, N., Pereira, F. C., & Bialek, W. (2000). The information bottleneck method. *ArXiv Preprint Physics/0004057*, 1–16. https://doi.org/10.1108/eb040537

Tsukahara, J. S., Harrison, T. L., & Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cognitive Psychology*, *91*, 109–123. https://doi.org/10.1016/j.cogpsych.2016.10.001

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

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2010). Bayesian Online Learning of the Hazard Rate in Change-Point Problems. *Neural Computation*, *22*(9), 2452–2476. https://doi.org/10.1162/NECO_a_00007

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zajkowski, W. K., Kossut, M., & Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *ELife*, *6*, 1–18. https://doi.org/10.7554/eLife.27430

Supplementary Figures

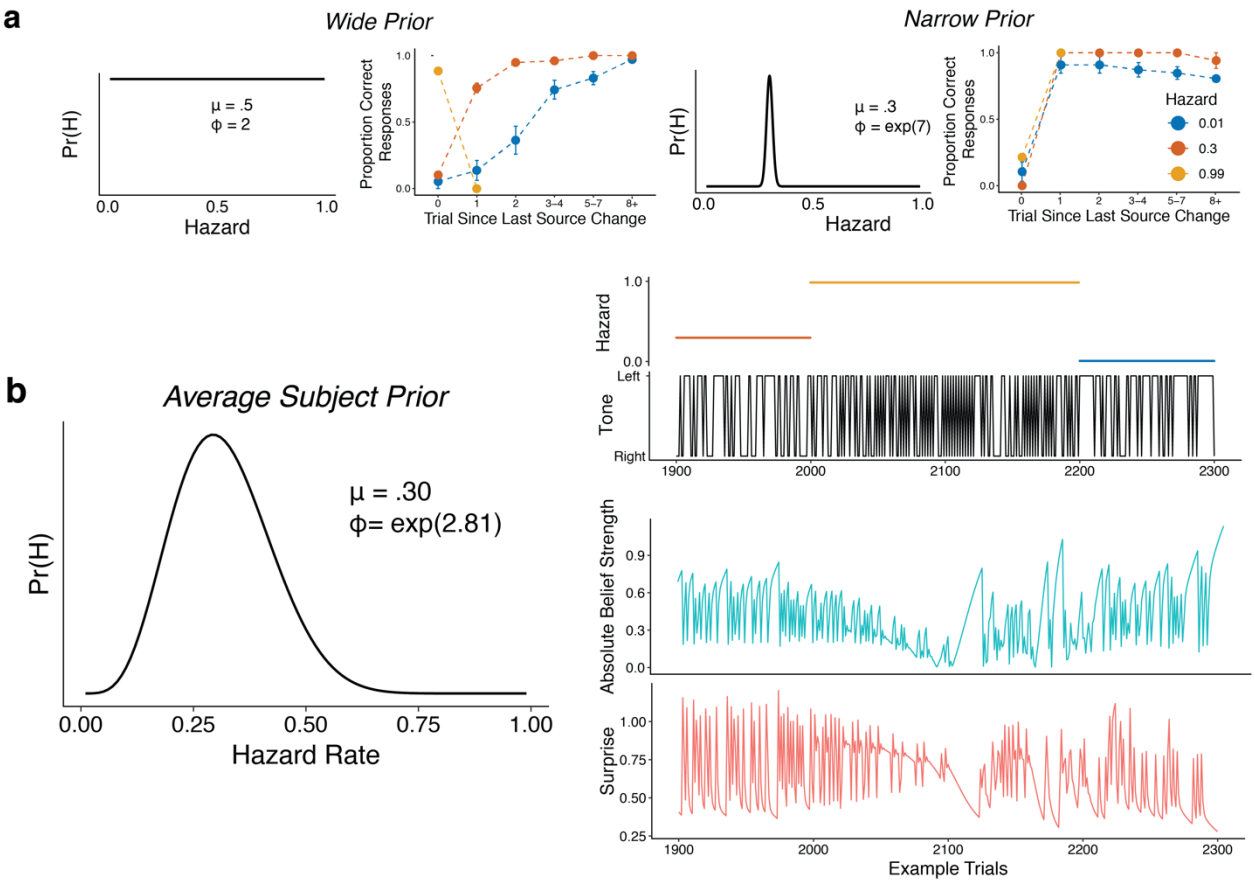

Figure S1. a) Example simulations from models with a wide prior (left panels) or a narrow prior (right panels) over hazard rate. For each pair of panels, the left panel shows the prior over hazard rate (governed by two parameters, as shown: mean μ and precision ϕ), and the right panel shows performance versus the number of trials since the previous change-point, separated by hazard rate (as in Fig. 1c,e). Points and error bars correspond to mean±SEM across simulated subjects. b) Average subject prior computed from fits of the Bayesian learning model to subject-specific behavioral data and the corresponding absolute belief strength (teal) and surprise values (pink) for 400 example trials (the tone location and hazard rate are shown above).

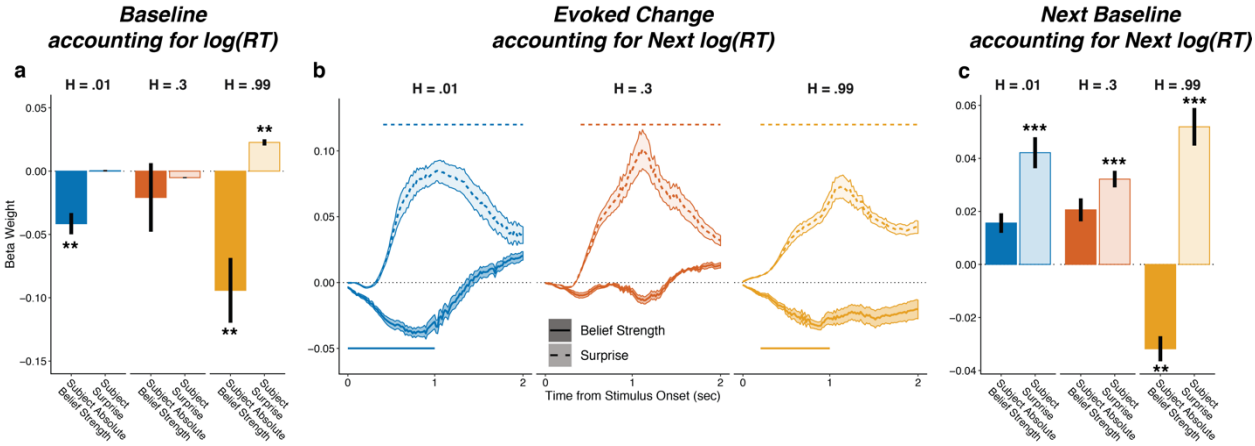

Figure S2. Pupils encode belief strength and surprise independently of RT. Linear mixed effects beta weights estimating the influence of subject absolute belief strength and surprise on: a) baseline pupil diameter b) evoked change, and c) next baseline pupil diameter after accounting for log RT (for baseline) or next log RT (for evoked change and next baseline). Both baseline pupil measures accounted for the previous evoked pupil change. The evoked change and next pupil baseline measures accounted for the previous baseline pupil and the evoked pupil change and baseline pupil on the current trial. Bars and errorbars in a and c or lines and shading in b correspond to mean±SEM across subjects. Straight lines above and below beta values in panel b indicate significant timepoints (FWE corrected $p<0.05$). In bar plots, **: $p<0.01$, ***: $p<0.001$.

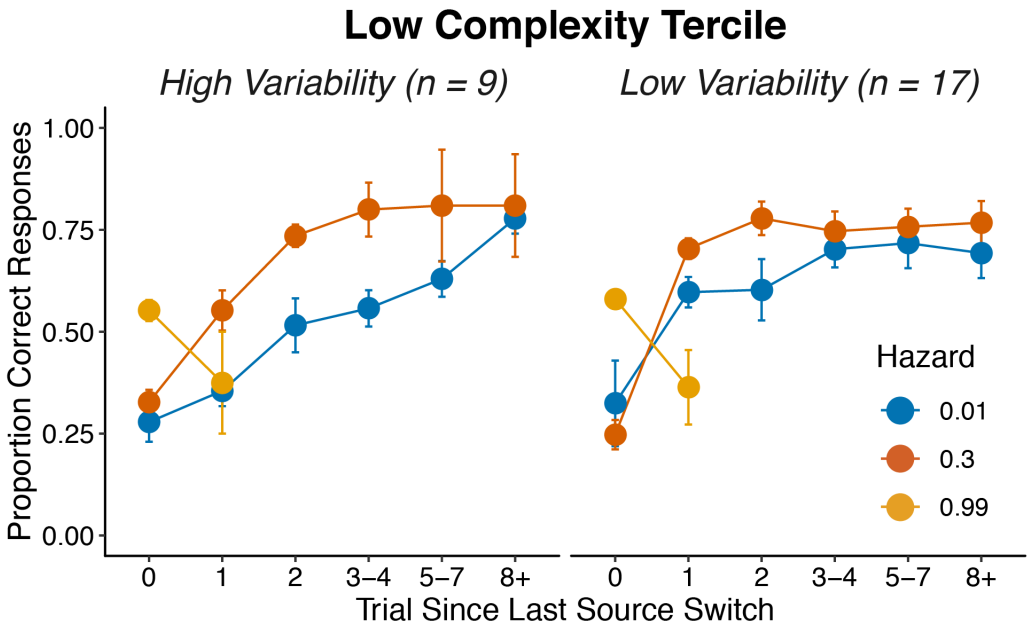

Figure S3. Mean proportion correct responses for high- and low-variability subjects in the low-complexity tercile. Low-variability subjects tended to use a strategy of predicting a source location based only on the previous tone location, which is similar to the strategy used by the Bayesian learning model with a narrow prior over hazard rate (Fig. 4a). High-variability subjects behaved in a more hazard-dependent manner, but with overall noisier predictions (i.e., proportion correct values closer to 0.5).