Humans adapt rationally to approximate estimates of uncertainty

Erdem Pulcu[a*] and Michael Browning[a,b]

[a] Department of Psychiatry, University of Oxford, Oxford, UK

[b] Oxford Health NHS Trust, Oxford, UK

*erdem.pulcu@psych.ox.ac.uk

Summary

Efficient learning requires estimation of, and adaptation to, different forms of uncertainty. If uncertainty is caused by randomness in outcomes (noise), observed events should have less influence on beliefs, whereas if uncertainty is caused by a change in the process being estimated (volatility) the influence of events should increase. Previous work has demonstrated that humans respond appropriately to changes in volatility, but there is less evidence of a rational response to noise. Here, we test adaptation to variable levels of volatility and noise in human participants, using choice behaviour and pupillometry as a measure of the central arousal system. We find that participants adapt as expected to changes in volatility, but not to changes in noise. Using a Bayesian observer model, we demonstrate that participants are, in fact, adapting to estimated noise, but that their estimates are imprecise, leading them to misattribute it as volatility and thus to respond inappropriately.

Keywords: Uncertainty, Bayesian Models, Pupillometry, Learning

Word count: 4313

Figures: 5

It is much easier to respond appropriately to an event if we know what has caused it. For example, if heavy traffic means that our drive into work takes longer than normal, the best course of action the next time we have to make the journey depends on what caused the traffic to be heavier (Yu & Dayan, 2005). If it was caused by a one-off or random event, such as a broken-down lorry, then we should continue using the same route as before, whereas if it was caused by some longer-term change, perhaps there are new road works nearby disrupting the traffic, we should consider a different route. Frequently, however, the causes of events are not obvious, we experience the heavy traffic but aren't sure why it has occurred. In these situations, the best we can do is make an educated guess, based on our experience, about what broad type of causal process has led to recent events. In the case of the drive into work, if the traffic has been heavier for a number of days in a row it is likely that some prolonged shift has occurred, and we should change routes, whereas if the traffic changes noisily from day to day, then we should probably stick with our usual route. In the learning literature, this problem is often framed as a competitive attribution of uncertainty to one of two types; expected uncertainty, which is caused by the variability of noisy associations and unexpected uncertainty, which is caused by longer lasting changes (sometimes called volatility) in an association (Behrens et al., 2007; Browning et al., 2015; Nassar et al., 2012; Pulcu & Browning, 2017; Yu & Dayan, 2005). The behavioural importance of this attribution process can be seen in the driving example given above; an event caused by noise requires the opposite behavioural response (continuing to use the same route) than the same event caused by volatility (switching routes). Consequently, effective decision making often depends on the accurate attribution of uncertainty, with misattribution having a substantial detrimental effect on choice (Pulcu & Browning, 2019).

The influence of events on subsequent choice can be estimated within a reinforcement learning framework as the learning rate parameter (Sutton & Barto, 2018), with a higher learning rate indicating a greater influence of the event on behaviour. As described above, the normative response to changes in volatility and noise is to use a higher learning rate when volatility is high and/or noise is low (Pulcu & Browning, 2019; Yu & Dayan, 2005). A large number of studies have found the predicted increase in learning rates in response to higher outcome volatility in human learners (Behrens et al., 2007, 2008; Browning et al., 2015; Gagne et al., 2020; Nassar et al., 2012; Pulcu & Browning, 2017). In contrast, the evidence for adaptation of learning in response to changes in outcome noise is less complete. Previous studies have described the expected reduction of learning rates when outcome noise is high, but only when the level of noise is explicitly signalled in a task (Diederen & Schultz, 2015), or when it is made unambiguous by virtue of being very much smaller than changes in outcome caused by volatility (Nassar et al., 2010, 2012). As illustrated in the driving example above, we are often faced with situations in which there exists significant

uncertainty about whether an event has been caused by volatility or noise. To date however, the degree to which human learners are able to discriminate between these types of uncertainty, when they are not explicitly labelled, has not been closely examined.

From a neurobiological perspective, activity of central modulatory neurotransmitter systems have been argued to represent distinct sources of uncertainty during learning, with central norepinepheric (NE)/locus coeruleus (LC) activity described as signalling changes in the associations (i.e. volatility) and central cholinergic activity representing noise (Yu & Dayan, 2005). Electrophysiological measures of LC activity in non-human primates have been shown to correlate with pupil dilation (Joshi et al., 2016) suggesting it may be possible to estimate activity in this system in humans using pupilometry. Taking this approach, indirect support for this role of the NE system has been provided by studies of human participants that report greater pupillary size in volatile relative to stable contexts (Browning et al., 2015; Nassar et al., 2012; Pulcu & Browning, 2017). However, the pupil also responds to other learning signals, such as surprise (Browning et al., 2015; O'Reilly et al., 2013; Preuschoff et al., 2011) and has been reported as being smaller when outcome noise is high (Nassar et al., 2012). Neuroimaging evidence suggests an association between activity in other central neurotransmitter nuclei, including the cholinergic basal forebrain, and pupil dilation (de Gee et al., 2017). Overall, this suggests that the pupillary signal may reflect a more general belief updating process (O'Reilly et al., 2013) rather than a specific volatility signal and thus that, like learning rates, pupil size should increase when noise is reduced as well as when volatility is increased.

In this paper we test whether human participants modify their learning in situations in which the attribution of uncertainty as volatility or noise is challenging (Figure 1a-c). We report the results of a study in which participants completed a learning task during which the noise and volatility of both win and loss outcomes were independently manipulated. Participant behaviour was characterised using learning rate parameters derived from reinforcement learning models of choice, while interpretation of the results was facilitated by a Bayesian Ideal Observer model that was developed to provide a benchmark comparator to participant behaviour (Behrens et al., 2007; Nassar et al., 2012; Piray & Daw, 2021; Pulcu et al., 2022) and by the collection of pupillometry data as a physiological marker of central neurotransmitter function (de Gee et al., 2017; Joshi et al., 2016). It was predicted that human participants would be able to adapt appropriately to the cause of the events they encountered—using a higher learning rate, and displaying increased pupil size, when volatility was high and when noise was low for both win and loss outcomes (Figure 1d).

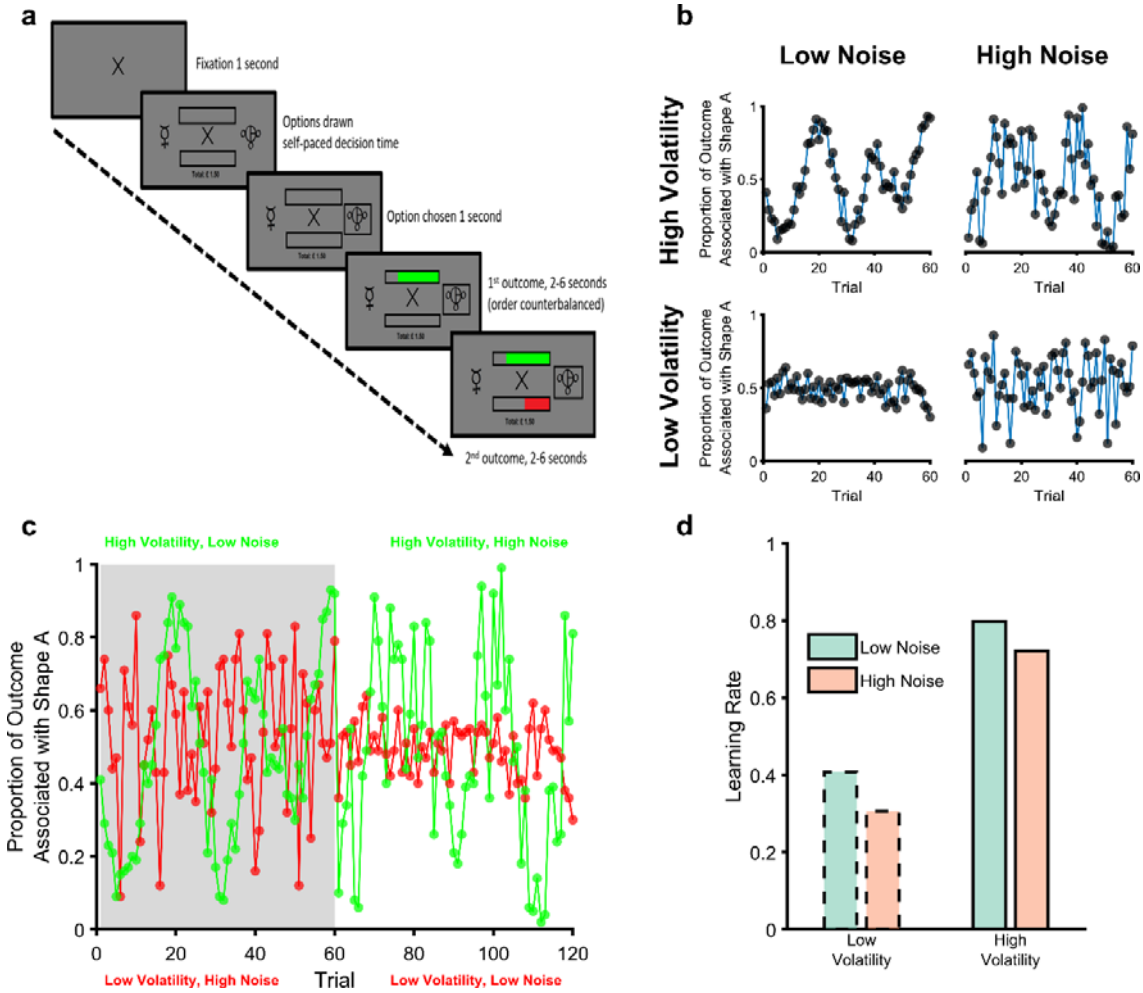

Figure 1. The Magnitude Learning Task (a) Timeline of one trial from the learning task. On each trial participants were presented with two abstract shapes and were asked to choose one of them. The empty bars above and below the fixation cross represented the total available wins and losses for the trial, the full length of each bar was equivalent to £1. Participants chose a shape and then were shown the proportion of each outcome that was associated with their chosen shape as coloured regions of the bars (green for wins and red for losses). The empty portions of the bars indicated the win and loss magnitudes associated with the unchosen option, allowing participants to infer which shape would have been the better option on every trial. The task consisted of six blocks of sixty trials each. The volatility and noise of the two outcomes varied independently between blocks with different shapes used in each block. Panel **b** illustrates outcomes from the four block types. As can be seen blocks with high volatility and low noise (top left), and those with low volatility and high noise (bottom right), present participants with a similar range of magnitudes. Participants therefore have to distinguish whether variability in the outcomes is caused by volatility or noise from the temporal structure of the outcomes rather than the size of changes in magnitude (cf; Diederen & Schultz, 2015; Krishnamurthy et al., 2017; Nassar et al., 2012). Panel **c** shows two example blocks (one block in grey, the other in white) with both win (green) and loss outcomes (red) displayed. Panel **d** shows the expected adaptation of learning rates in response to the manipulation of volatility and noise; for both win and loss outcomes, learning rates should be increased when volatility is high and when noise is low.

Results

Participant Demographics

70 participants (see Supplementary Table 1 for demographic information) completed a learning task in which they had to choose one of two stimuli based on the separately estimated magnitudes of win and loss outcomes associated with the stimuli (Figure 1). Participants were able to learn the best option to choose in the task, selecting the most highly rewarded option on an average of 71% of trials (range 65% - 74%).

Experimental Manipulation of Volatility and Noise Influences Participant Choice Behaviour

As explained above, high levels of volatility and low levels of noise should increase the degree to which outcomes influence choice behaviour. A crude metric of this effect is provided by examining participant choice as a function of the previous outcome. In the task, a win outcome of >50p or a loss outcome of <50p associated with Shape A prompts participants to select Shape A in the subsequent trial, with the other outcomes (i.e. win <50p and loss >50p) prompting choice of Shape B. The influence of the outcomes on choice can therefore be roughly estimated as the relative proportion of trials in which Shape A was chosen when it was prompted by a previous outcome of a given magnitude, compared to when Shape B was prompted. Analysis of this choice metric (Figure 2a-b) found the expected effect of volatility, with participant choice being more influenced by previous outcomes when volatility was higher ($F(1,696)=99.8$, $p<0.001$). An effect of noise was observed, but in the opposite direction to expected, with outcomes influencing choice more when noise was increased ($F(1,696)=4.79$, $p=0.03$). No significant difference between the influence of win and loss outcomes was found ($F(1,696)=1$, $p=0.32$) and there was no interaction between volatility and noise ($F(1,693)=0.61$, $p=0.4$). Having found some evidence of an impact of the uncertainty manipulations on a crude measure of subject choice we next sought to characterise this effect using reinforcement learning models fitted to participant choices.

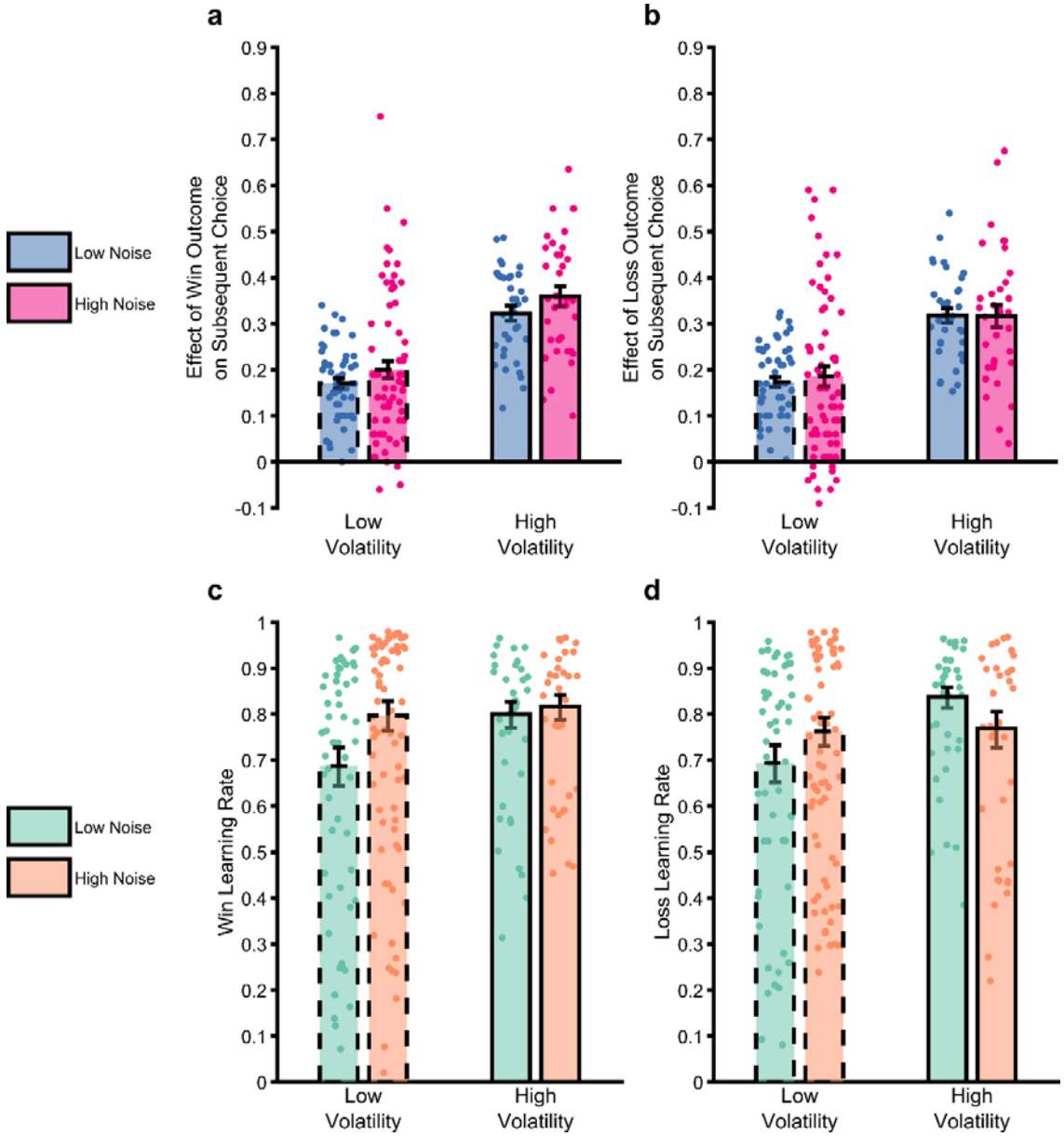

Figure 2. The impact of uncertainty manipulations on participant choice. Panels **a** and **b** report a summary metric for the effect of win and loss outcomes on subsequent choice. The metric was calculated as the proportion of trials in which an outcome of magnitude 51-65 associated with Shape A was followed by choice of the shape prompted by the outcome (i.e. Shape A for win outcomes, Shape B for loss outcomes) relative to when the outcome magnitude was 49-35 (see methods and materials for more details). We focused on this outcome range as these range of magnitudes were covered by all volatility x noise conditions and it was dictated by the relatively smaller range coverage in the low volatility low noise condition (also see Figure 1C loss outcomes shown in red between trials 60-120). The higher this number, the greater the tendency for a participant to choose the shape prompted by an outcome. As can be seen, the outcome of previous trials had a greater influence on participant choice when volatility was high, with a small effect of noise, in the opposite direction to that predicted. Panels **c** and **d** report the win and loss learning rates estimated from the same data. Again, the expected effect of volatility is observed, this time with no consistent effect of noise. Bars represent the mean (±SEM) of the data, with individual data points superimposed.

157 **Participants Adjust Normatively to Changes in Volatility but not Noise**

158 We aimed to capture the computational process that underlies participant choice behaviour by
159 fitting different reinforcement learning models to choice data separately for each block of the task
160 and each participant. The best fitting RL model included separate learning rates for win and loss
161 outcomes allowing estimation of the degree to which participants adjusted these learning rates in
162 response to the block-wise changes in outcome volatility and noise (see supplementary Materials
163 and Methods for model comparison and selection analyses).

164 Consistent with the analysis of choice data reported above, there was a significant main effect of
165 volatility (Figure 2c-d; $F(1,696)=22.2$, $p<0.001$), with a higher learning rate used when volatility was
166 high. There was no main effect of noise ($F(1,696)=0.63$, $p=0.43$) on learning rate or outcome valence
167 ($F(1,696)=0.15$, $p=0.7$). An interaction between volatility and noise ($F(1,693)=7.74$, $p=0.006$) was also
168 significant. A higher volatility led to a significantly raised learning rate when noise was low
169 ($F(1,383)=27.1$, $p<0.01$) , with a non-significant increase when noise was high ($F(1,311)=1.13$,
170 $p=0.29$). Higher noise was associated with a non-significant reduction in learning rates when
171 volatility was high ($F(1,347)=2.57$, $p=0.11$) but to a significant increase in learning rate when volatility
172 was low ($F(1,347)=4.7$, $p=0.031$).

173 In summary, analysis of both crude choice data and learning rates indicates that participants
174 adapted appropriately to changes in the volatility of learned associations but did not show a
175 consistent response to changes in noise. In the next section we utilise a Bayesian Observer Model
176 (BOM) to investigate potential causes for this relative insensitivity to noise.

177

178 **Using a Bayesian Observer Model to Characterise Noise Insensitivity**

179 Bayesian Observer Models (BOM) can be used as normative benchmarks against which human
180 behaviour may be compared (Behrens et al., 2007; Nassar et al., 2012; Piray & Daw, 2021; Pulcu et
181 al., 2022). BOMs are generally not fit to participant choice, rather these models invert a generative
182 process assumed to underlie observed events and provide an estimate of the belief of an idealised
183 agent exposed to the same outcomes as participants. We developed a BOM (Pulcu et al., 2022)
184 based on the generative process underlying the outcome magnitudes of our task (Figure 3a). The
185 BOM explicitly estimates the volatility and noise of the outcomes and uses these estimates to
186 influence its belief about the likely magnitude of upcoming outcomes (see methods for more
187 details). We first tested whether the BOM reproduced the normative learning rate adaptation to
188 changes in volatility and noise described in the introduction, by exposing the model to the same

outcomes as participants and using the model's belief about the likely magnitude of the win and loss outcome on each trial to generate choices. We then estimated the effective learning rate of the model by fitting the same RL model used to analyse participants' choices to the model's choices. These learning rates are presented in Figure 3f (Figure 3e reproduces the learning rates of participants, averaged across wins and losses, for comparison). As can be seen the BOM adapts as expected, using a higher learning rate both when volatility increases ($F(1,696)=422$, $p<0.001$) and when noise decreases ($F(1,696)=21.2$, $p<0.001$). No effect of outcome valence or interaction between volatility and noise (all $p>0.09$) was observed.

Having shown that an optimal learner adjusts its learning rate to changes in volatility and noise as expected, we next sought to understand the relative noise insensitivity of participants. In these analyses we "lesion" the BOM, to reduce its performance in some way, and then assess whether doing so recapitulates the pattern of learning rate adaptation observed for participants (Fig 3e). In other words, we damage the model so it performs less well and then assess whether this damage makes the behaviour of the BOM (shown in Fig 3f) more closely resemble that seen in participants (Fig 3e). First, we tested the impact of completely removing the ability of the BOM to adjust to changes in either volatility (Figure 3b) or noise (Figure 3c) by removing the top nodes of the model (i.e. kmu or vs respectively). Removing these nodes forces the BOM to estimate the mean volatility or noise across all task blocks rather than estimating local periods where they are higher or lower (see supplementary video). As illustrated in Figure 3g-h, neither of these lesions recapitulates the pattern of learning rates observed in participants, with the volatility lesioned model attributing increased volatility to noise and thus decreasing its learning rate during periods of higher volatility (main effect of volatility; $F(1,696)=11.9$, $p<0.001$) and the SD-lesioned model treating any form of uncertainty as volatility and thus increasing its learning rate in response to increased noise (main effect of noise; $F(1,696)=227$, $p<0.001$). This suggests that human participants are able to adapt to changes in outcome volatility and noise to some degree but are less sensitive to these changes than the intact BOM.

We next assessed whether a relative degrading of the model's representation of volatility and noise (Figure 3d) altered its behaviour in a manner similar to participants. This was achieved by independently coarsening the model's representation of volatility and noise, with the degree of coarsening selected to make the model's choices as similar as possible to those of a given participant. Details of this coarsening process are provided in the methods section, but in simple terms, at one extreme, the intact model's beliefs about current volatility and noise are represented as probability distributions over many possible values, with the number of values used gradually reduced during coarsening, until the coarsest model treats each form of uncertainty as being either

223 “high” or “low”. As can be seen from Figure 3i, this relative degrading of the model’s representation
224 of uncertainty more closely recapitulated the learning rates observed in participants, with a
225 significant increase in learning rate in response to increased volatility ($F(1,696)=59$, $p<0.001$) and no
226 effect of noise ($F(1,696)=2.3$, $p=0.13$). In total the BOM fitted to participant choices had 5
227 parameters (i.e. volatility and SD acuity for rewards and losses and a single inverse temperature
228 term). This was compared with two reinforcement learning models: the measurement model
229 described previously, which was fitted to individual blocks and therefore had many more parameters
230 (18 in total; learning rates for wins and losses for each block, one inverse temperature term per
231 block), and a simple version of the same model which was fitted across all blocks and had 3
232 parameters in total (win and loss LR and one inverse temperature parameter). Model comparison
233 between the BOM and the RW models based on BIC scores favoured the BOM (mean (SD) for BOM;
234 235 (54), for complex RL measurement model; 281 (57), for simple RL model; 246 (58)).

235 In the next sections we characterise how coarsening the BOM changes its behaviour and assess
236 whether it provides an accurate account of participants’ noise insensitivity.

237

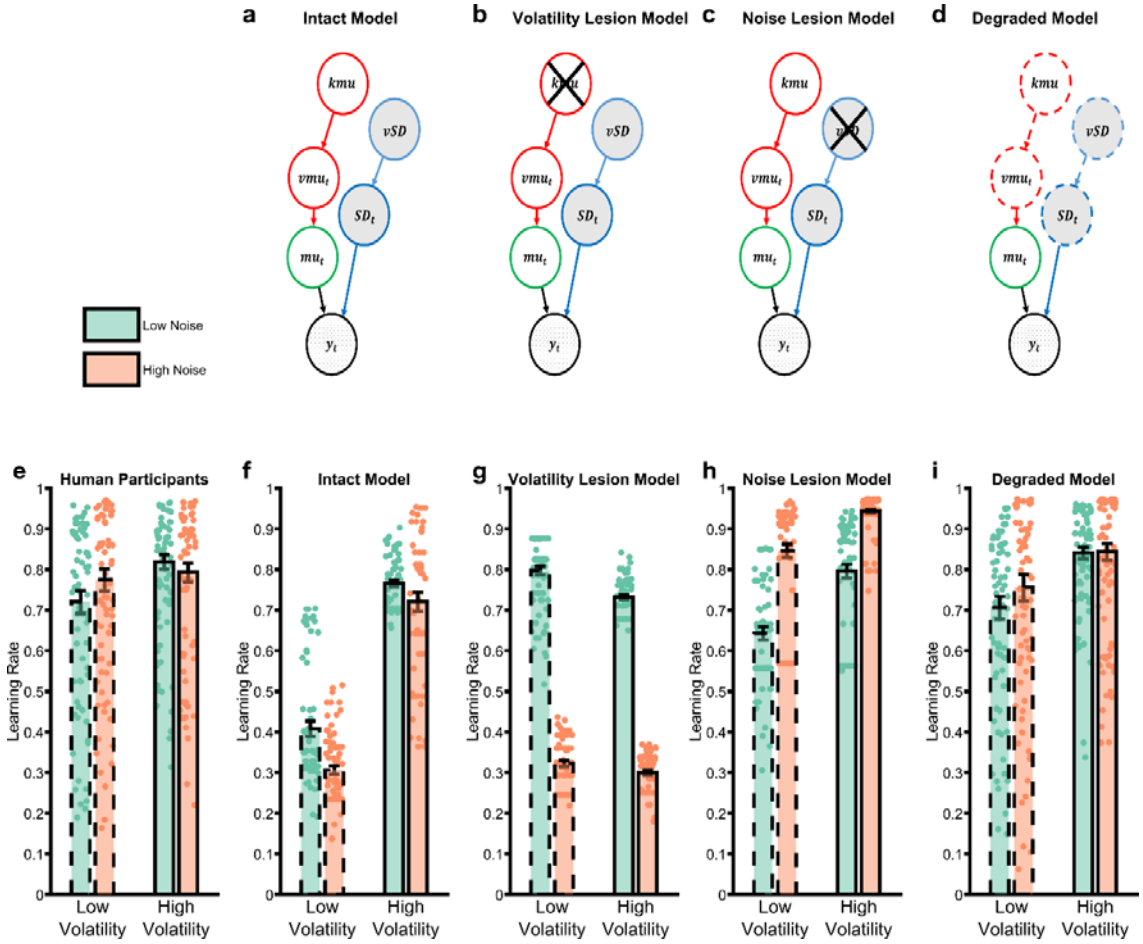

Figure 3. The Behaviour of Bayesian Observer Models. Bayesian Observer Models (BOM) invert generative descriptions of a process, indicating how an idealised observer may learn. We developed a BOM based on the generative model of the task we used (**a**). Details of the BOM are provided in the methods, briefly it assumes that observations (y_i) are generated from a Gaussian distribution with a mean (mu_i) and standard deviation (SD_i). Between observations, the mean changes with the rate of change controlled by the volatility parameter (vmu_i). The standard deviation and volatility of this model estimate the noise and volatility described for the task. The last parameters control the change in volatility (kmu) and standard deviation (vs) between observations, allowing the model to account for different periods when these types of uncertainty are high and others when they are low. The BOM adjusts it learning rate in a normative fashion (**f**), increasing it when volatility is higher, or noise is lower. The BOM was lesioned in a number of different ways in an attempt to recapitulate the learning rate adaptation observed in participants (shown in panel **e**). Removing the ability of the BOM to adapt to changes in volatility (**b**) or noise (**c**) did not achieve this goal (**g**, **h**). However, degrading the BOMs representation of uncertainty (**d**) was able to recapitulate the behavioural pattern of participants (**i**). Bars represent the mean (±SEM) of participant learning rates, with raw data points presented as circles behind each bar.

The Degraded BOM Misattributes Noise as Volatility

The BOM was degraded by reducing the number of bins it used to represent volatility and/or noise, until its behaviour most closely matched that of participants. This process led to a greater coarsening of the noise than the volatility dimension (Figure 4a; $F(1,69)=49$, $p<0.001$), with no effect of outcome valence ($F(1,69)=0.73$, $p=0.4$), suggesting that the degraded model maintained a generally less

precise representation of noise than volatility. In order to investigate the impact of this coarsening on the model's beliefs, we used the degraded BOM's estimates of volatility and noise to categorise task trials as either high or low volatility/noise (i.e. trials in which the model's estimates of these variables were higher/lower than the mean) and compared these to the same trial labels generated by the intact BOM. Consistent with the greater degradation of the noise dimension, coarsening the model caused it to miscategorise more trials which the intact BOM had labelled as having high than low noise (Figure 4b; $F(1,69)=30.7$, $p<0.01$) with no effect of volatility ($F(1,69)=1.9$, $p=0.17$) or outcome valence ($F(1,69)=0.004$, $p=0.95$). As illustrated in Figure 4c, when the degraded BOM miscategorised high noise trials, it tended to label them as having high, rather than low, volatility. Overall, these results indicate that coarsening the BOM caused it, relative to the intact BOM, to misattribute high noise trials as high volatility trials.

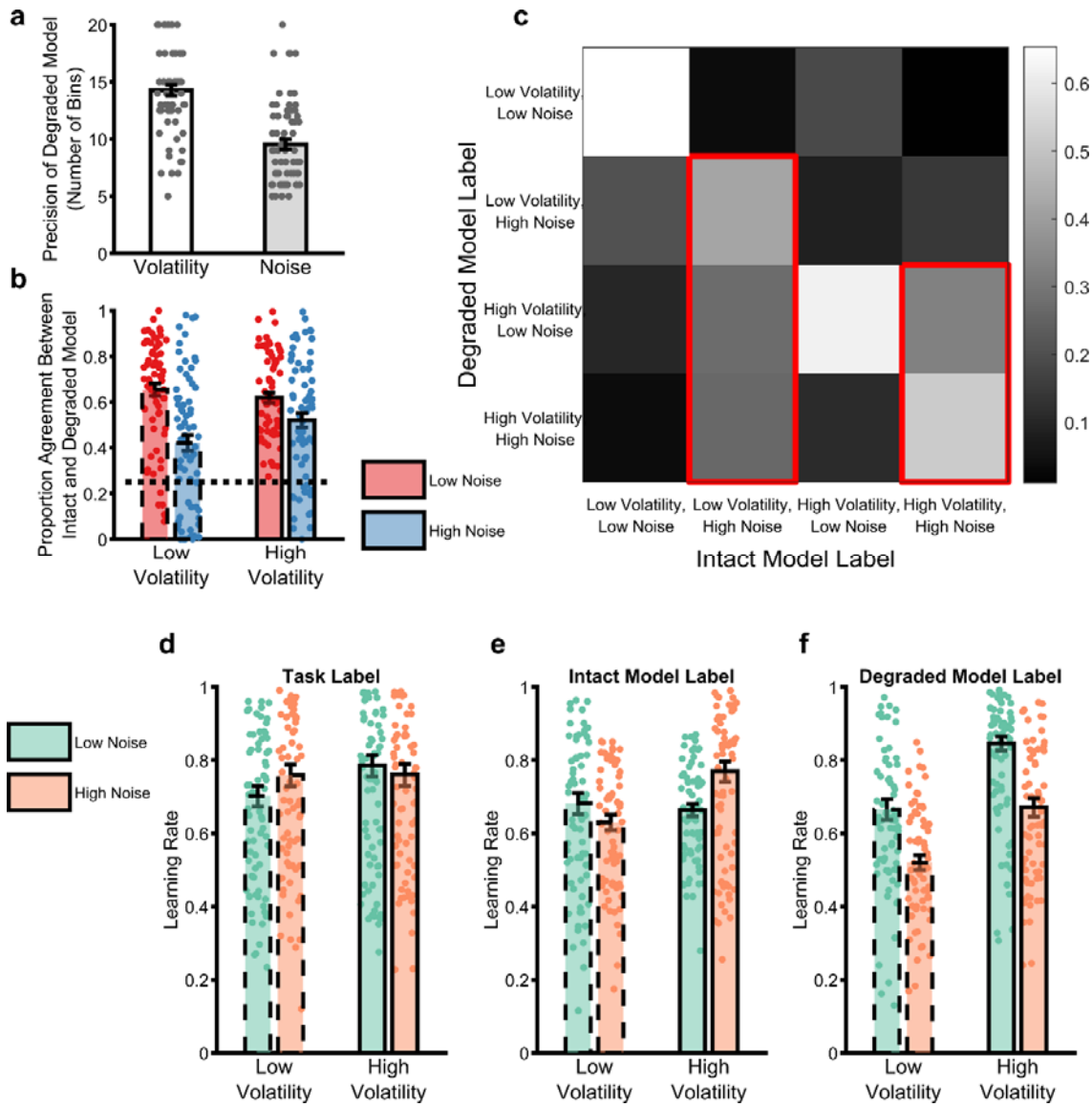

Figure 4. Analysis of the behaviour of the degraded BOM. The process of degrading the BOM involved reducing the number of bins used to represent the volatility and noise dimensions independently until the choice of the model matched that of participants. Panel **a** illustrates the number of bins selected by this process for the volatility and noise dimensions (averaged across win and loss outcomes). As can be seen the degraded BOM maintained a less precise representation of noise than volatility. In order to understand the behaviour of the degraded model, the model's estimated vmu_i and SD_i were used to label individual trials as high/low volatility and noise (NB greater than or less than the mean value of the estimates). These trial labels were compared with the same labels from the intact model, which were used as an ideal comparator (panels **b** and **c**). Panel **b** illustrates the proportion of trials in which the labels of the two models agreed, arranged by the ground truth labels of the full model and averaged across win and loss outcomes. The dotted line indicates the agreement expected by chance. The degraded model trial labels differed from those of the full model particularly for high noise trials, with no impact of trial volatility. Panel **c** provides more details on how the degraded model misattributes trials. In this figure, the labels assigned by the full model are arranged along the x axis. The colour of each square represents the proportion of trials with a specific full model label that received the indicated label of the degraded model (arranged along the y axis). The diagonal squares illustrate agreement between models as reported in panel **b**. As highlighted by the red outlines, trials which the full model labelled as having high noise were generally mislabelled by the degraded model as having high volatility. Reanalysis of participant choices using the trial labels provided by the full (panel **e**) and degraded (panel **f**) models indicate that participants adapt their learning rates in a normative fashion when the degraded model trial labels are used (panel **f**), but not when the full model labels are used (panel **e**). Panel **d** illustrates the same analysis using the original task block labels for comparison. Bars represent the mean (±SEM) of participant learning rates, with raw data points presented as circles behind each bar. See supplementary figure 1 for a comparison of the behaviour of the degraded BOM with an alternative fitted model.

The Degraded BOM Rescues Optimal Behaviour

The process of fitting the degraded BOM to participant behaviour can be understood as searching for a configuration of the model (akin to a grid-based maximum likelihood estimation) in which participant choice conforms to the normative response to volatility and noise coded in the model's structure. In other words, participants' learning rates should increase when the degraded BOM's estimate of volatility is high and, critically, when it estimates that noise is low. We demonstrate this by reanalysing participant behaviour, using the trial labels of the degraded BOM to indicate periods of low/high volatility and noise in place of the task block labels used in the original analysis. In effect, this approach allows us to test whether participants make internally consistent errors during learning, i.e. reducing their learning rates for outcomes that they thought — instead of the actual task structure — were associated with high volatility and/or low noise. As can be seen (Figure 4f), participants significantly increased their learning rate when the degraded BOM estimated volatility to be high ($F(1,566)=86$, $p<0.001$) and noise to be low ($F(1,566)=81$, $p<0.001$). In control analyses, this normative response to uncertainty was not seen when the labels from the intact rather than the degraded BOM were used (Figure 4e), or when the BOM's representation of outcome mean was degraded, rather than its estimates of volatility and noise (supplementary materials).

Assuming Human Participants Use the Degraded BOM's Estimates of Volatility and Noise also Rescues Normative Pupillary Response

If the degraded BOM is a fair representation of how participants are performing the learning task, then we would expect it to be better able to explain physiological markers of uncertainty estimation than the simple task block structure or the intact BOM. Specifically, participants' pupils should be larger when the degraded BOM thinks that volatility is high and when it thinks noise is low. We first show (Figures 5a-c) that participants' pupils do not adapt normatively to the task block structure, with no main effect of block volatility ($F(1,1723)=0.002$, $p=0.9$) and an increase of pupil size in response to higher noise ($F(1,1723)=13.8$ $p<0.001$). In contrast, analysis using the trial labels derived from the degraded model (Figure 5d-f) recovered the expected increase in pupil size in response to both raised volatility ($F(1,2067)=105$, $p<0.001$) and reduced noise ($F(1,2067)=42.3$, $p<0.001$) suggesting that the model provides a reasonable measure of participants' estimates of these parameters. Finally, we tested whether the degraded BOM was able to explain more variance in the pupil data than the intact BOM. In order to do this, we first regressed participants' pupil data against the estimated volatility and noise of the intact BOM, as well as a range of other task related factors (Figure 5g; see methods for more details of analysis). Having removed the variance accounted for by these factors, we then regressed the residuals of this first level analyses against the degraded model's estimates of volatility and noise. This second level analysis (Figure 5h-i) indicated that the degraded model was able to account for variance associated with outcome noise that was not explained by the full model ($F(1,286)=4.1$, $p=0.04$), but did not explain additional variance associated with outcome volatility ($F(1,286)=0.1$, $p=0.75$). In summary, assuming that participants used the degraded BOM's estimates of outcome volatility and noise rescued the normative pattern of physiological adaptation during the task.

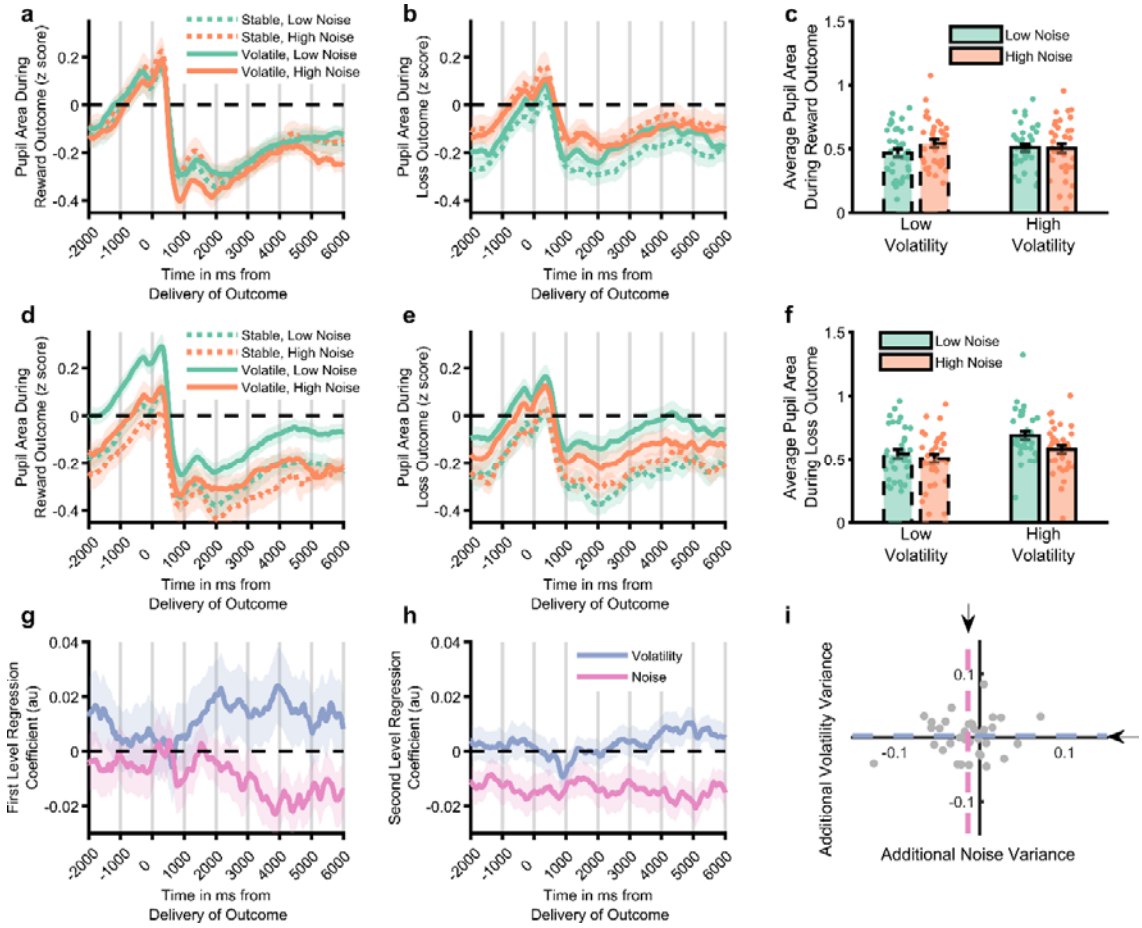

Figure 5: Analysis of pupillometry data. Z-scored pupil area from 2 seconds before to 6 seconds after win (panel **a**) and loss (panel **b**) outcomes, split by task block. Lines illustrate average size, with shaded area illustrating SEM. Panel **c** Pupil size averaged across whole outcome period and both win and loss outcomes. Pupil size did not systematically vary by task block. Panels **d-f**, as above but using the trial labels derived from the degraded model. Pupil size was significantly larger for trials labelled as having high vs. low volatility and low vs. high noise. Panel **g** displays the mean (SEM) effect of volatility and noise as estimated by the full BOM derived from a regression analysis of pupil data. The residuals from this analysis were then regressed against the estimated volatility and noise from the degraded model. A time course of the regression weights from this analysis is shown in panel **h**, with the mean coefficients across the whole period shown in panel **i**. The degraded model's estimated noise accounted for a significant amount of variance not captured by the full model (pink line in **h** is below 0, the mean effect across the period is represented by dashed lines and arrows in panel **i**). See supplementary figure 2 for comparison of the degraded BOM with an alternative fitted model.

Discussion

Humans respond in a rational, if approximate, manner to the causal statistics of dynamic environments. We found that participants adapted as expected to changes in outcome volatility, but were relatively insensitive to changes in noise. Using a degraded Bayesian Observer Model (BOM) to

characterise participants' behaviour suggested that they responded appropriately to a relatively coarse estimation of the level of noise, that led to its misattribution as volatility. Analysis of pupillometry data using the degraded model again suggested that participants were responding normatively to changes in estimated noise, but that these estimates diverged from the true noise of experienced outcomes. These results illustrate that human learners are able adapt to the statistical properties of their environment, but during this process they make internally consistent errors, utilising higher learning rates as a result of misattributing environmental noise as volatility which also leads to suboptimal choice.

Using a task in which volatility and noise varied independently between blocks, we found that human learners adapted as expected (Behrens et al., 2007; Browning et al., 2015; Nassar et al., 2012; Pulcu & Browning, 2019) to blockwise changes in the volatility of both win and loss outcomes, increasing the learning rate used when volatility was high vs. low. In contrast, the expected reduction of learning rates in response to increased outcome noise was not apparent, with participants employing a significantly higher learning rate in response to increased noise when volatility was low and a numerically lower learning rate when volatility was high. The absence of a normative response to blockwise changes in noise is at odds with previous work which has described a reduction in learning rates during periods of high noise (Diederen & Schultz, 2015; Nassar et al., 2010, 2012). However, in this previous work the level of noise was either explicitly presented to participants (as a bar on screen representing the standard deviation of the generative process in Diederen & Schultz) or was made unambiguous by being very different from changes caused by volatility (in Nassar et al, noise was generated using an SD of 5 or 10, while the average change due to volatility was 100). By design, in the current task high noise and volatility resulted in a similar range of magnitudes (Figure 1b) forcing participants to use the temporal sequence of outcomes to discriminate between the different forms of uncertainty. Our behavioural results suggest that, in the absence of unambiguous differences between outcomes caused by volatility and those caused by noise, participants' ability to estimate and/or adapt to changes in noise is reduced. Interestingly, a recent study reported that participants do not adjust their choice or estimated confidence in response to variability in the orientation of arrays of visual gratings (Herce Castañón et al., 2019), suggesting that an insensitivity to outcome noise may be a general feature of human decision making, rather than a specific component of learning.

Noise fundamentally limits the reliability of information (MacKay, 2003) and ignoring it has a clear detrimental impact on inference (Figure 3h), causing agents to be unnecessarily influenced by chance events (Pulcu & Browning, 2019). It would therefore be surprising if human learners were completely insensitive to this process, particularly given evidence that they can respond normatively

386 when the level of noise is unambiguous (Diederen & Schultz, 2015; Nassar et al., 2010, 2012). We
387 developed an ideal Bayesian Observer Model (BOM; Behrens et al., 2007; Nassar et al., 2010; Piray &
388 Daw, 2021; Pulcu et al., 2022) to investigate the degree to which participants were adapting to
389 noise. The intact BOM displayed the expected behavioural response to changes in both volatility and
390 noise (Figure 3f) and, as a result, did not accurately capture the behaviour of participants (Figure 3e).
391 Completely removing the BOM's ability to adapt to noise (or volatility) did not recapitulate
392 participant choice behaviour (Figure 3g-h), whereas coarsening its representation of volatility and
393 noise, produced a much closer match (Figure 3i). This suggests that participants were relatively,
394 rather than completely insensitive to noise and that they tended to misattribute high noise as
395 volatility (Figure 4). However, an important caveat to this interpretation is that the degree of
396 coarsening was selected using participants' choices. The better behavioural match of the coarsened
397 BOM to participant learning rates may therefore be simply because this model was fitted to the
398 same choices used to calculate the learning rates, whereas the intact and fully lesioned models were
399 not. We therefore sought to validate the coarsened BOM by assessing its ability to account for
400 participants' pupillary data, and by comparing it with an alternative fitted BOM which coarsened the
401 representation of the generative mean, rather than the estimated uncertainty (see supplementary
402 materials). Participants' pupil size did not vary systematically between different block types, whereas
403 they were significantly larger when the degraded BOM estimated volatility to be high and noise to
404 be low (Figure 5a-f). Similarly, the estimated noise of the degraded BOM accounted for additional
405 variance in pupil size, over and above the intact BOM (Figure 5g-i). In contrast, the alternative mean-
406 degraded BOM did not recapitulate participants' learning rates (supplementary figure 3) and was not
407 able to account for changes in participant pupil size (supplementary figure 4). The finding that
408 participants' pupil size covaries in the expected direction with the degraded BOM's estimated levels
409 of both volatility and noise provides some reassurance that the model is capturing the dynamics of
410 participants' uncertainty estimates. More generally, the presence of both volatility and noise signals
411 in this data, indicate that, as suggested previously (Nassar et al., 2012; O'Reilly et al., 2013), the
412 pupillometry signal reflects general belief updating rather than specifically volatility.

413 An outstanding question is why participants might be particularly insensitive to changes in outcome
414 noise. It is tempting to try to answer this question by reference to the processes by which the BOM
415 was coarsened (i.e. the insensitivity was caused by a reduction in the precision by which noise was
416 represented in a multi-dimensional probability distribution). However, the BOM described here was
417 developed as an algorithmic description of how the learning task may be solved. As far as we are
418 aware, there is little evidence that it accurately describes the cognitive or neural implementation of
419 uncertainty estimation. Alternative algorithmic approaches to the general problem of uncertainty

estimation have been described (Kalman, 1960; Nassar et al., 2010; Piray & Daw, 2021; Pulcu & Browning, 2019), including simpler approaches that avoid computationally expensive representations of multi-dimensional distributions (Kalman, 1960; Nassar et al., 2010) and which therefore may be more likely implementational candidates. In other words, the current results indicate that human learners are relatively insensitive to changes in outcome noise, but do not specify the lower level mechanisms that determine this effect.

Previous work examining the neural representations of uncertainty have tended to report correlations between brain activity and some task-based estimate of one form of uncertainty at a time (Behrens et al., 2007; Walker et al., 2020, 2023). We are not aware of work that has, for example, systematically varied volatility and noise and reported distinct correlations for each. An interesting possibility as to how different forms of uncertainty may be encoded is suggested by parallels with the neuronal decoding literature. One question addressed by this literature is how the brain decodes changes in the world from the distributed, noisy neural responses to those changes, with a particular focus on the influence of different forms of between-neuron correlation (Averbeck et al., 2006; Kohn et al., 2016). Specifically, signal-correlation, the degree to which different neurons represent similar external quantities (required to track volatility) is distinguished from, and often limited by, noise-correlation, the degree to which the activity of different neurons covaries independently of these external quantities. One possibility relevant to the current study, which resembles the underlying logic of the BOM, is that a population of neurons represents the estimated mean of the generative process that produces task outcomes. In this case, volatility would be tracked as the signal-correlation across this population, whereas noise would be analogous to the noise-correlation and, crucially, misestimation of noise as volatility might arise as misestimation of these two forms of correlation. While the current study clearly cannot adjudicate on the neural representation of these processes, our finding of distinct behavioural and physiological responses to the two forms of uncertainty, does suggest that separable neural representations of uncertainty are maintained.

A related question is whether other, non-Bayesian model formulations may be able to account for participants' learning adaptation in response to volatility and noise. Of note, the reinforcement learning model used to measure learning rates in separate blocks does not achieve this goal—as this model is fitted separately to each block rather than adapting between blocks (NB the simple reinforcement learning model that is fitted across all blocks does not capture participant behaviour, see supplementary information). One candidate class of model that has potential here is latent-state models (Cochran & Cisler, 2019), in which the variance and unexpected changes in the process being learned (which have a degree of similarity with noise and volatility respectively) is estimated and

used to alter the model's rates of updating as well as the estimated number of states being considered. Using the model described by Cochran and Cisler, we were unable to replicate the learning rate adaptation demonstrated by participants in the current study (see supplementary information) although it remains possible that other latent state formulations may be more successful. In conclusion, human learners adapt rationally, to estimates of the volatility and noise of experienced outcomes. However, these estimates are approximate leading to a relative insensitivity to outcome noise.

Methods

Experimental model and subject details

Participants. 70 English-speaking participants aged between 18 and 65 were recruited from the general public using print and online advertisements. A previous study (Pulcu & Browning, 2017) on behavioural response to changes in volatility reported an effect size of *d*=0.7. As the effect size of a noise manipulation was not clear, we recruited a sample size sufficient to detect an effect size of half this value (*d*=0.35) with 80% power. Participants were excluded from the study if they had any psychological or neurological disorders or were currently on psychotropic medication. No exclusion criteria related to task performance were used.

Method details

General procedure. Participants attended a single study visit during which they completed the learning task. The study was approved by the University of Oxford Central Research Ethics Committee (R49753/RE001). All participants provided written informed consent to take part in the study, in accordance with the Declaration of Helsinki.

Behavioural Paradigm. The reinforcement learning (RL) task consisted of six blocks, each comprising 60 trials. In each trial, participants were presented with two abstract shapes taken from the Agathodaimon font (i.e. shape A and shape B). Two different shapes were used in each block, with rest sessions between blocks. The shapes were presented randomly on either side of the screen. Participants were explicitly instructed that this randomised location did not influence the outcome magnitudes. Participants attempted to accumulate as much money as possible by learning the likely magnitude of the wins and losses associated with each shape and using this information to guide their choice. On each trial, participants chose one of two shapes, with their choice highlighted by a black frame (see Fig 1a). Following the choice, the win and loss amounts associated with the chosen shape were presented, in randomised order, for a jittered period (2-6 sec, mean: 4 sec) inside two

487 empty bars, above and below the fixation cross. The win amount was shown as a green area in the
488 upper bar, and the loss amount represented as a red area in the lower bar. The total length of each
489 bar represented £1 (i.e. of wins or losses) and thus the amount associated with the chosen shape
490 was the proportion of the bar filled by the green/red areas (e.g. three quarters of the upper bar
491 being green, would mean that the chosen option was associated with a win of 75p). Participants
492 were informed that the unshaded area of each bar was the amount associated with the unchosen
493 option. Thus, on each trial participants knew how much they had won/lost and how much they
494 would have won/lost if they had chosen the other option. This feature simplified the task; rather
495 than having to separately estimate the wins and losses associated with each shape, participants only
496 had to estimate these values for one shape (with the other shape being £1 minus this value). For
497 each trial participants received the difference between the win and loss amounts associated with
498 their choice. A running total amount of money was displayed in the centre of the screen, under the
499 bars and was updated at the beginning of the subsequent trial with the recent winnings. Participants
500 were informed that the task would be split into 6 blocks, that they had to learn which was the best
501 option to choose, and that this option may change over time. They were not informed about the
502 different forms of uncertainty we were investigating or of the underlying structure of the task (that
503 uncertainty varied between blocks).

504 The wins and the losses associated with each shape followed independent outcome schedules
505 (Figure 1b), generated from a Gaussian distribution. In each block, the win and loss outcomes had
506 either high or low volatility and high or low noise. When volatility was low, the mean of the Gaussian
507 distribution remained constant, when volatility was high the mean changed from between 25-40 and
508 60-75 every 9-15 trials. When noise was low the standard deviation of the Gaussian was set to 5,
509 whereas when noise was high the standard deviation was 35. As can be seen from Figure 1b, these
510 schedules resulted in similar ranges of outcome magnitudes for periods of high noise and high
511 volatility. The first block for every participant had high volatility and low noise for both win and loss
512 outcomes and was used to familiarise participants with the task. Choices from this block were not

513 used in the analyses presented (although including them does not alter the reported pattern of
514 results). The schedules in the remaining five blocks were presented in a randomised order with the
515 constraint that, across both win and loss outcomes, each of the four combinations of volatility and
516 noise level (Figure 1B) were presented either 2 or 3 times. Thus, while each participant completed at
517 least two blocks with each of the four combinations of high/low volatility/noise, the specific pairings
518 of win and loss volatility/noise levels, differed across participants. This approach was used in
519 preference to a fully factorial design in order to keep the total task duration to a manageable level.
520 At the end of the experiment, participants were paid one fifth of their total winnings, plus a £15
521 baseline rate for turning up to take part.

522 Pupillometry data was collected for 36 of the 70 participants. During collection of pupillary data, the
523 task was presented on a VGA monitor connected to a laptop computer running Presentation
524 software version 18.3 (Neurobehavioural Systems). An identical behavioural version of the task,
525 presented using Psychtoolbox 3.0 on MATLAB (MathWorks Inc.), was used to collect behavioural
526 data from the remaining 34 participants. In the pupillometry version, participants' heads were
527 stabilised using a head-and-chin rest placed 70 cm from the screen on which the eye tracking system
528 was mounted (Eyelink 1000 Plus; SR Research). The eye tracking device was configured to record the
529 coordinates of both of the eyes and pupil area at a rate of 500 Hz. The task stimuli were drawn on
530 either side of a fixation cross which marked the middle of the screen and were offset by 7° visual
531 angle. The testing session lasted approximately 70 min per participant.

532 **Analysis of Choice Data**

533 *Non-model based measure of the influence of outcomes.* The manipulation of uncertainty in the
534 reinforcement learning task is expected to alter the degree to which participants' choices are
535 influenced by the outcomes they experience. A simple, if somewhat crude, measure of this influence
536 can be calculated as the proportion of trials in a block in which participants select the choice
537 prompted by the win or loss outcomes on the previous trial. Generally, win outcomes of >50p and

538 loss outcomes of <50p associated with a shape will prompt selection of the same shape on the next
539 trial, whereas other outcomes will prompt selection of the alternative shape. The overall effect of
540 win outcomes on choice can therefore be estimated as:

$$P_{(choice==A\,|\,previous\ win\ outcome\ for\ A>50p)} - P_{(choice==A\,|previous\ win\ outcome\ for\ A<50p)}$$

541 That is, the probability of choosing shape A, given that, on the previous trial, a win of >50p was
542 associated with shape A – the probability of choosing Shape A, given that, on the previous trial a win
543 of <50p was associated with Shape A. Similarly, the effect of loss outcomes is estimated as:

$$P_{(choice==A\,|\,previous\ loss\ outcome\ for\ A<50p)} - P_{(choice==A\,|previous\ loss\ outcome\ for\ A>50p)}$$

544 However, choice is also influenced by the magnitude of the outcome; a win of 90p will have a
545 greater effect on subsequent choice than a win of 55p. Blocks with high levels of either volatility or
546 noise have more extreme magnitudes than blocks with low levels of both (Figure 1b) which will bias
547 any comparison of this metric between blocks. In order to limit the effect of this bias, we estimated
548 the simple choice metric only for trials in which the previous outcome lay in the range of magnitudes
549 common to all four blocks, 35-65.

550 *Reinforcement Learning Model*: While the choice metric described above provides a relatively
551 transparent measure of the influence of task outcomes on choice, it does not account for differences
552 in outcome magnitude making it liable to bias. We therefore fitted a simple reinforcement learning
553 model to measure block-wise learning rates, which provide a more principled estimate of the degree
554 to which choices are influenced by outcomes. The model combines a learning phase in which the
555 magnitude of wins and losses associated with a shape are estimated (note that it is not necessary to
556 learn the magnitudes associated with the other shape, as these are simply 1- those described below)

$$Qwin_a_{(t+1)} = Qwin_a_{(t)} + \alpha_{win}(win_{(t)} - Q_{win_a(t)})$$

$$Qloss_a_{(t+1)} = Qloss_a_{(t)} + \alpha_{loss}(loss_{(t)} - Q_{loss_a(t)})$$

557 In these equations, $Qwin_a_{(t)}$ and $Qloss_a_{(t)}$ are the estimated win and loss magnitudes associated
558 with Shape A on trial *t*, $win_{(t)}$ and $loss_{(t)}$ are the observed win and loss outcome magnitudes and
559 α_{win} and α_{loss} are the win and loss learning rates. These values are then combined in a decision
560 phase such that:

$$Pchoice_a_{(t)} = \frac{1}{1 + e^{-\beta(Q_{win_a(t)} - Q_{loss_a(t)})}}$$

561 Where $Pchoice_a_{(t)}$ is the probability that Shape A will be chosen on trial *t* and β is a single inverse
562 decision temperature. This model was initiated with $Qwin_a_{(0)}$ = $Qloss_a_{(0)}$ = 0.5 and the three free
563 parameters ($win_{(t)}$, $loss_{(t)}$ and β) were estimated for each block and each participant by calculating
564 the joint posterior probability given participant choice, marginalising each parameter and deriving
565 the parameters' expected values (Behrens et al., 2007; Browning et al., 2015). See supplementary
566 materials for model selection data.

567 Some analyses reported in the paper (i.e. where trials are labelled as high/low volatility and high/low
568 noise by the Bayesian Observer Model rather than by task block) cannot be modelled using this
569 block-wise approach (as different types of trial are interleaved throughout the task, rather than
570 blocked). In these analyses a similar, single model was fit across all trials in the task. This model had
571 8 different learning rates (separate win and loss learning rates, for each combination of high/low
572 volatility and high/low noise labelled trials) and a single inverse temperature parameter. Although
573 this model is somewhat less flexible than the blockwise modelling approach (i.e. it has 8, rather than
574 10 learning rates, and 1 rather than 5 inverse temperatures), it produces the same pattern of results
575 when applied to participant choices split by task block (all estimated learning rates correlate at
576 *r*>0.8, Figure 2c-d show results from blockwise fitting, Figure 3e from the simpler model). This
577 simpler model was fit using stan, with 5000 burn in and 5000 estimation trials, with posterior
578 convergence visually checked and rhat values of less than 1.1 accepted.

579 Note that neither of these models describe how participants adjust to different levels of volatility
580 and noise, they simply estimate the learning rates used in each block/type of trial, which are
581 expected to vary in response to differences in levels of uncertainty (in contrast, the Bayesian
582 Observer Model described below does estimate uncertainty and adjust to levels of uncertainty).

583 *Bayesian Observer Model:* A recursive, grid-based Bayesian Observer Model (BOM) was developed,
584 similar to that described by Behrens and colleagues (Behrens et al., 2007; Pulcu et al., 2022). The
585 BOM is based on a generative process (see Figure 3), and described fully in Pulcu et al. (Pulcu et al.,
586 2022). Below we summarise the key aspects of the model.

587 The BOM assumes that the observed outcomes at a given time point t, y_t, are generated from a
588 Gaussian distribution with an unknown mean, μ_t, and standard deviation, e^{SD_t}, with the later
589 producing noise in the observed outcomes (Figure 1b-c).

590 $$y_t \sim N(\mu_t, e^{SD_t})$$

591 As illustrated in Figure 1b-c, the mean of this distribution may change between time points, leading
592 to volatility in the task environment, with this change described by a second level Gaussian
593 distribution, centered on the current mean and with a standard deviation of e^{vmu_t}. The mean of the
594 generative Gaussian distribution in the following trial is drawn from:

595 $$P(\mu_{t+1}) \sim N(\mu_t, e^{vmu_t})$$

596 Both the noise (SD_t) and volatility (vmu_t) parameters can also change between time points with
597 their change governed by Gaussian distributions centered on their current value with standard
598 deviations of e^{vSD} and e^{kmu} respectively. These higher-level parameters allow the model to account
599 for periods in which noise and volatility are high and other periods in which they are low (for
600 example, as caused by the uncertainty changes between task blocks).

$$P(vmu_{t+1}) \sim N(vmu_t, e^{kmu})$$

$$\mathrm{P}(SD_{t+1}) \sim N(SD_t, e^{vSD})$$

The BOM estimates the joint posterior probability of the five causal parameters, given the choice outcome it has observed. The joint probability distribution at time point t is defined as:

$$P(joint_t) = P(mu, vmu, kmu, SD, vSD \mid y_{t-1}, y_{t-2}, \dots, y_1)$$

This joint probability distribution can be thought of as the BOM's belief about the values of each parameter in the generative model. A Markovian assumption (i.e. that nodes of the model are sufficient to describe the generative process) simplifies this process and illustrates the recursive update performed by the BOM:

$$P(joint_t) = P(mu_t, vmu_t, kmu_t, SD_t, vSD_t \mid joint_{t-1}, y_{t-1})$$

We initialized the joint posterior, before observation of any task outcomes as a uniform distribution. The BOM performs the update, first using Bayes' rule to incorporate the effect of the most recently observed outcome, and then accounts for the drifting parameters by using the conditional probability of the new value of the drifting parameter, given the initial value and drift rate (See; Pulcu et al., 2022 for a detailed account of this updating process):

$$p(joint_t \mid joint_{t-1}, y_{t-1}) =$$

$$\iiint p(joint_{t-1} \mid y_{t-1}) p(SD_t \mid SD_{t-1}, vSD)\, p(vmu_t \mid vmu_{t-1}, kmu) \dots$$

$$p(mu_t \mid mu_{t-1}, vmu_t), dSD_{t-1}, dvmu_{t-1}, dmu_{t-1}$$

The value of each node is derived at every time point by marginalizing over all but the relevant dimension of the joint probability distribution and calculating the expected value of that dimension.

During the task, the shapes presented to participants change between each task block, which means that, at the start of each block, participants have to relearn the mean associated with each shape. This was dealt with in the BOM by flattening the mu dimension of the joint probability distribution at

621 the start of each trial (i.e. replacing the values of the mean dimension, with the average of the joint
622 distribution across this dimension). The effect of this is to reset the model's belief about the actual
623 magnitude associated with the two new shapes, while maintaining its belief about the overall
624 volatility and noise of the outcomes.

625 The BOM was provided with the win and loss outcomes (as values between 0 and 1) for each trial,
626 across all trials in the task (excluding the first practice block, although including this did not alter the
627 pattern of results). It treated the two outcomes as independent (i.e. the win outcome did not
628 influence estimates for the loss outcome and vice versa) and transformed the outcomes to the
629 infinite real line using the logistic transform before estimating the posterior probability (Pulcu et al.,
630 2022).

631

632 *Lesioning the Bayesian Observer Model*: A number of different lesions were applied to the BOM.
633 First, it's ability to estimate changes in either volatility or noise was removed. This was achieved
634 simply by removing the *kmu* or *vSD* nodes from the BOM (reducing the dimensionality of the joint
635 distribution by one in each case). The effect of this is to force the BOM to estimate the mean
636 volatility and noise (respectively) across the whole task, rather than to modify its estimates of these
637 parameters between trials.

638 The second approach induced a graded, rather than absolute, lesion. This was achieved by reducing
639 the precision with which the BOM represented the volatility-related nodes (*vmu* and *kmu*) and/or
640 the noise related nodes (*SD* and *vSD*). More specifically, the BOM's estimates of the values of each of
641 the five nodes are encoded on a five dimensional grid, with each dimension on the grid representing
642 the possible range of values of a particular node, from low to high, using a fixed number of points.
643 The probability ascribed by the model to a specific point on this dimension is the relative probability
644 that the value of the node lies within the bin of values that is closer to the point, than to adjacent
645 points. For example, say the value of volatility (*vmu*) ranged from 0 to 10 and was represented by 10

646 bins. In this case volatility would be represented by a probability mass function over the 10 bins
647 (<0.5, 0.5-1.5, 1.5-2.5, …, > 9.5). Lesioning occurred by independently varying the number of bins
648 used in the volatility-related and/or noise-related dimensions, from a maximum of 20, to a minimum
649 of 2 (i.e. with only 2 bins volatility/noise would be represented as simply "high" or "low"). The
650 degree of lesioning selected for each individual participant was determined as the number of bins
651 for the volatility and noise dimensions that, after passing the model estimates through a softmax
652 action selector with a single inverse temperature parameter (i.e. as described for the RL model),
653 maximized the likelihood that the model would make the same choices as the participant, across all
654 task blocks. This process of lesioning therefore progressively coarsens the BOM's representation of
655 the two types of uncertainty and selects the degree of coarsening that results in choices as similar as
656 possible to participants (see supplementary materials for an alternative model that coarsens the
657 representation of the mean values).

658

659 **Pupilometry Data Preprocessing.** Pupilometry data were collected using the Eyelink II system
660 (SRresearch) from both eyes, sampled at 500Hz. Preprocessing involved the following steps: Eye
661 blinks were identified using the built in filter of the Eyelink system and were removed from the data.
662 A linear interpolation was implemented for all missing data points (including blinks). The resulting
663 trace was subjected to a low pass Butterworth filter (cut-off of 3.75 Hz), z transformed across the
664 session (Browning et al., 2015; Nassar et al., 2012), and then averaged across the two eyes. The pupil
665 response to the win and the loss outcomes were extracted separately from each trial, using a time
666 window based on the presentation of the outcomes. This included a 2-s pre-outcome period, and a
667 6-s period following outcome presentation. Individual trials were excluded from the pupilometry
668 analysis if more than 50% of the data from the outcome period had been interpolated (mean =6.7%
669 of trials) (Browning et al., 2015). The first 5 trials from each block were not used in the analysis as
670 initial pupil adaption can occur in response to luminance changes in this period (Browning et al.,

671 2015; Nassar et al., 2012). The preprocessing resulted in two sets of timeseries per participant, one
672 set containing pupil size data for each included trial when the win outcomes were displayed and the
673 other when the loss outcomes were displayed. These pupil area data were binned into one second
674 bins across the outcome period for analysis (NB Figure 5a-f). This analysis was supplemented by an
675 individual regression approach (Figure 5g-i) in which individual participants' pupil area timeseries
676 was first regressed against estimated trialwise volatility and noise from the intact BOM (Figure 5g),
677 as well as a number of control variables (constant term, amount won/lost on trial (i.e. magnitude of
678 outcome), valence of outcome (win or loss), order in which outcomes were presented (win first/loss
679 first), trial number (1:360), whether shape chosen switched on next trial or not (1:0)). The residuals
680 from this regression were then regressed against estimated trialwise volatility and noise from the
681 degraded BOM (Figure 5h,i). These regression analyses resulted in timeseries of beta-weights that
682 were analysed in the same manner as raw pupil size data.

683 **Quantification and statistical analysis**

684 Behavioural data were analysed using linear mixed effect models (*fitlme* function of Matlab (2022a))
685 with participant ID included as a random factor and volatility, noise and valence added as fixed
686 factors. Two-way interactions between fixed effects were also tested (main effects are reported
687 from models without interaction terms). Addition of random slopes for any of the fixed factors
688 decreased LME model fit statistics and so were not included (Matuschek et al., 2017). Analysis of
689 timeseries pupillometry data included the additional fixed effect factor of time across the outcome
690 period. Learning rates were transformed to the infinite real line using a logistic transform before
691 analyses (untransformed data are displayed in figures for ease of interpretation). The normality of
692 the distribution of the residuals of the LME analyses were checked both visually and with a one-
693 sample Kolmogorov-Smirnov test. Changes in the classification of trials between the full and
694 degraded BOM (Figure 4b) were analysed using a repeated measures ANOVA with within subject
695 factors of volatility, noise and valence. Raw data are superimposed on all summary figures.

Code and Data Availability

Study data and analysis scripts, including code for the various models used are available at: https://osf.io/j7md3/.

Acknowledgements

We would like to thank James Gunnell for help in collecting the data. This study was funded by a MRC Clinician Scientist Fellowship awarded to MB (MR/N008103/1). MB was supported by the Oxford Health NIHR Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Author Contributions

MB and EP conceived the study, EP collected the data, MB and EP wrote the paper.

Disclosures

MB has received travel expenses from Lundbeck for attending conferences and consultancy from Jansen, CHDR and Novartis. EP declares no potential conflict of interest.

References

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*(5), 358–366. https://doi.org/10.1038/nrn1888

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

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

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

Cochran, A. L., & Cisler, J. M. (2019). A flexible and generalizable model of online latent-state learning. *PLoS Computational Biology*, *15*(9), e1007331. https://doi.org/10.1371/journal.pcbi.1007331

de Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T., Nieuwenhuis, S., & Donner, T. H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. *eLife*, *6*, e23232. https://doi.org/10.7554/eLife.23232

Diederen, K. M. J., & Schultz, W. (2015). Scaling prediction errors to reward variability benefits error-driven learning in humans. *Journal of Neurophysiology*, *114*(3), 1628–1640. https://doi.org/10.1152/jn.00483.2015

Gagne, C., Zika, O., Dayan, P., & Bishop, S. J. (2020). Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife*, *9*, e61387. https://doi.org/10.7554/eLife.61387

Herce Castañón, S., Moran, R., Ding, J., Egner, T., Bang, D., & Summerfield, C. (2019). Human noise blindness drives suboptimal cognitive inference. *Nature Communications*, *10*(1), 1719. https://doi.org/10.1038/s41467-019-09330-7

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

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problem. *Transactions of the ASME*, *82*(D), 35–45.

Kohn, A., Coen-Cagli, R., Kanitscheider, I., & Pouget, A. (2016). Correlations and Neuronal Population Information. *Annual Review of Neuroscience*, *39*, 237–256. https://doi.org/10.1146/annurev-neuro-070815-013851

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

MacKay, D. J. C. (2003). *InformationTheory,Inference,and Learning Algorithms* (1st ed.). Cambridge University Press.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

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. *Nat Neurosci*, *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. *Journal of Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

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), E3660-3669. https://doi.org/10.1073/pnas.1305373110

Piray, P., & Daw, N. D. (2021). A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications*, *12*(1), 6587. https://doi.org/10.1038/s41467-021-26731-9

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*, 115. https://doi.org/10.3389/fnins.2011.00115

Pulcu, E., & Browning, M. (2017). Affective bias as a rational response to the statistics of rewards and punishments. *eLife*, *6*. https://doi.org/10.7554/eLife.27879

Pulcu, E., & Browning, M. (2019). The Misestimation of Uncertainty in Affective Disorders. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2019.07.007

Pulcu, E., Saunders, K. E. A., Harmer, C. J., Harrison, P. J., Goodwin, G. M., Geddes, J. R., & Browning, M. (2022). Using a generative model of affect to characterize affective variability and its response to treatment in bipolar disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(28), e2202983119. https://doi.org/10.1073/pnas.2202983119

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introdcution* (Second). MIT Press.

Walker, E. Y., Cotton, R. J., Ma, W. J., & Tolias, A. S. (2020). A neural basis of probabilistic computation in visual cortex. *Nature Neuroscience*, *23*(1), 122–129. https://doi.org/10.1038/s41593-019-0554-5

Walker, E. Y., Pohl, S., Denison, R. N., Barack, D. L., Lee, J., Block, N., Ma, W. J., & Meyniel, F. (2023). Studying the neural representations of uncertainty. *Nature Neuroscience*, *26*(11), 1857–1867. https://doi.org/10.1038/s41593-023-01444-y

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