Lapses in perceptual judgments reflect exploration

Sashank Pisupati*[1,2], Lital Chartarifsky-Lynn*[1,2], Anup Khanal[1] & Anne K. Churchland[1]

[1] *Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA*

[2] *Watson School of Biological Sciences, Cold Spring Harbor, New York, USA*

ABSTRACT

During perceptual decision making, subjects often display a constant rate of errors independent of evidence strength, referred to as "lapses". Their proper treatment is crucial for accurate estimation of perceptual parameters, however they are often treated as a nuisance arising from motor errors or inattention. Here, we propose that lapses can instead reflect a dynamic form of exploration. We demonstrate that perceptual uncertainty modulates the probability of lapses both across and within modalities on a multisensory discrimination task in rats. These effects cannot be accounted for by inattention or motor error, however they are concisely explained by uncertainty-guided exploration. We confirm the predictions of the exploration model by showing that changing the magnitude or probability of reward associated with one of the decisions selectively affects the lapses associated with that decision in uncertain conditions, while leaving "sure-bet" decisions unchanged, as predicted by the model. Finally, we demonstrate that muscimol inactivations of secondary motor cortex and posterior striatum affect lapses asymmetrically across modalities. The inactivations can be captured by a devaluation of actions corresponding to the inactivated side, and do not affect "sure-bet" decisions. Together, our results suggest that far from being a nuisance, lapses are informative about subjects' action values and deficits thereof during perceptual decisions.

INTRODUCTION

Perceptual judgments are often modeled using noisy ideal observers (e.g., Signal detection theory, Green, Swets, et al., 1966; Bayesian decision theory, Dayan and Daw, 2008) that explain subjects' errors as a consequence of noise in sensory evidence. This predicts an error rate that decreases with increasing sensory evidence, capturing the sigmoidal relationship often seen between evidence strength and subjects' decision probabilities (i.e. the psychometric function; Fig. 1).

Human and non-human subjects often deviate from these predictions, displaying an additional constant rate of errors independent of the evidence strength known as "lapses", leading to errors even on extreme stimulus levels (Wichmann and Hill, 2001; Gold and Ding, 2013; Carandini and Churchland, 2013; Busse et al., 2011). Despite the knowledge that ignoring or improperly fitting lapses can lead to serious mis-estimation of psychometric parameters (Wichmann and Hill, 2001; Prins and Kingdom, 2018), the cognitive mechanisms underlying lapses remain poorly understood. A number of possible sources of noise have been proposed to explain lapses, typically peripheral to the decision-making process.

One class of explanations for lapses relies on pre-decision noise added due to fluctuating attention, which is often operationalized as a small fraction of trials on which the subject fails to attend to the stimulus (Wichmann and Hill, 2001). On these trials, it is assumed that the subject cannot specify the stimulus (i.e. sensory noise with infinite variance, Bays, Catalao, and Husain, 2009) and hence guesses randomly or in proportion to prior beliefs. This model can be thought of as a limiting case of the Variable Precision model, which assumes that fluctuating attention has a more

2

graded effect of scaling the sensory noise variance (Garrido, Dolan, and Sahani, 2011), giving rise to heavy tailed estimate distributions, resembling lapses in the limit of high variability (Shen and Ma, 2019; Zhou et al., 2018). Temporal forms of inattention have also been proposed to give rise to lapses, where the animal ignores early or late parts of the evidence (impulsive or leaky integration, Erlich et al., 2015).

An alternative class of explanations for lapses relies on a fixed amount of noise added after a decision has been made, commonly referred to as "post-categorization" noise (Erlich et al., 2015) or decision noise (Law and Gold, 2009). Such noise could arise from errors in motor execution (e.g. finger errors , Wichmann and Hill, 2001), non-stationarities in the decision rule arising from computational imprecision (Findling et al., 2018), suboptimal weighting of choice or outcome history (Roy et al., 2018; Busse et al., 2011) or random variability added for the purpose of exploration (eg."ϵ-greedy" decision rules).

A number of recent observations have cast doubt on fixed early- or late-stage noise as satisfactory explanations for lapses. For instance, many of these explanations predict that lapses should occur at a constant rate, while in reality, lapses are known to reduce in frequency with training in non-human primates (Law and Gold, 2009). Further, they can occur with different frequencies for different stimuli even within the same subject (in rodents, Nikbakht et al., 2018; and humans, Mihali et al., 2018; Bertolini et al., 2015), suggesting that they may reflect task-specific, associative processes that can vary within a subject.

Lapse frequencies are even more variable across subjects and can depend on the subject's

age and state of brain function. For instance, lapses are significantly higher in children and patient populations than healthy adult humans (Witton, Talcott, and Henning, 2017). Moreover, a number of recent studies in rodents have found that perturbing neural activity in secondary motor cortex (Erlich et al., 2015) and striatum (Yartsev et al., 2018; Guo et al., 2018) has dramatic, asymmetric effects on lapses in auditory decision-making tasks. Because these perturbations were made in structures known to be involved in action selection, an intriguing possibility is that lapses reflect an integral part of the decision-making process, rather than a peripheral source of noise. However, because these studies only tested auditory stimuli, they did not afford the opportunity to distinguish sensory modality-specific deficits from general decision-related deficits. Taken together, these observations point to the need for a deeper understanding of lapses that accounts for effects of stimulus set, learning, age and neural perturbations.

Here, we leverage a multisensory task in rodents to reveal a novel explanation for lapses: uncertainty-guided exploration, a well known strategy for balancing exploration and exploitation in value-based decisions (Dayan and Daw, 2008). This is also known as Thompson sampling and can be operationalized as a dynamic "softmax" decision rule (Gershman, 2018). We confirm the predictions of the exploration model by manipulating reward under conditions of varying uncertainty. Finally, we demonstrate that suppressing secondary motor cortex or posterior striatum unilaterally has an asymmetric effect on lapses that generalizes across sensory modalities, but only in uncertain conditions. This can be accounted for by an action value deficit on the inactivated side, reconciling the proposed perceptual and value-related roles of these areas and suggesting that lapses are informative about the subjective values of actions.

4

RESULTS

Testing ideal observer predictions in perceptual decision-making

We leveraged an established decision-making task (Raposo, Sheppard, et al., 2012; Raposo, Kaufman, and Churchland, 2014; Sheppard, Raposo, and Churchland, 2013; Licata et al., 2017) (Fig. 1) in which freely moving rats judge whether the fluctuating rate of a 1000 ms series of auditory clicks and/or visual flashes (rate range: 9 - 16 Hz) is high or low compared with an abstract category boundary (Fig. 1a - c). Using Bayesian decision theory, we constructed an ideal observer for our task that selects choices that maximize expected reward (See Methods: Modelling). To test whether behavior matches ideal observer predictions, we presented multisensory trials with matched visual and auditory rates (i.e., both modalities carried the same number of events/s; Fig. 1c, bottom) interleaved with visual-only or auditory-only trials. This allowed us to separately estimate the sensory noise in the animals' visual and auditory system, and compare the measured performance on multisensory trials to the predictions of the ideal observer.

Performance was assessed using a psychometric curve, i.e., the probability of high-rate decisions as a function of stimulus rate (Fig. 1f). The ideal observer model predicts a relationship between the slope of the psychometric curve and the noise in the animal's estimate: the higher the standard deviation (σ) of sensory noise, the more uncertain the animals estimate of the rate and the shallower the psychometric curve. On multisensory trials, the ideal observer should have a more certain estimate of the rate, driving a steeper psychometric curve (Fig. 1e, visual [blue] and auditory [green] σ values are larger than multisensory σ [red] and Fig. 1f, red curve is steeper than

5

102 green and blue curves). Since this model does not take lapses into account, it would predict perfect

103 performance on the easiest stimuli regardless of uncertainty, and thus all curves should asymptote at

104 0 and 1 (Fig 1f).
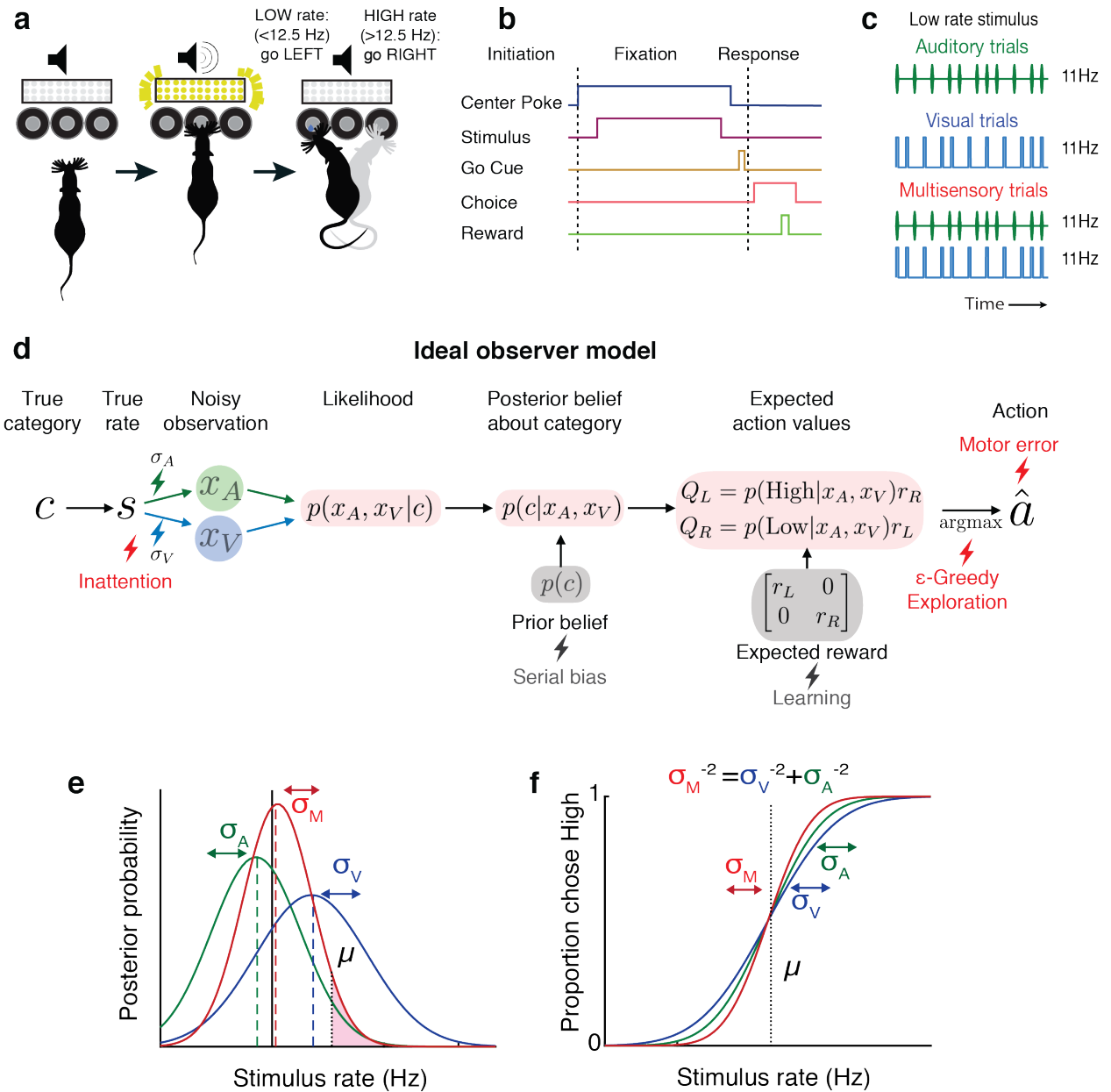
106 **Figure 1** **Testing ideal observer predictions in perceptual decision-making**. **(a)** Schematic drawing of

107 rate discrimination task. Rats initiate trials by poking into a center port. Trials consist of visual stimuli

6

presented via a panel of diffused LEDs, auditory stimuli presented via a centrally positioned speaker or

multisensory stimuli presented from both. Rats are rewarded with a 24μl drop of water for reporting high

rate stimuli (greater than 12.5 Hz) by poking in the right port and low rate stimuli (lower than 12.5 Hz) with

leftward choices. **(b)** Timeline of task events. **(c)** Example stimulus on auditory (top), visual (middle) and

multisensory trials (bottom). Stimuli consist of a stream of events separated by long (100 ms) or short (50

ms) intervals. Multisensory stimuli consist of visual and auditory streams carrying the same underlying

rate. Visual, auditory and multisensory trials were randomly interleaved (40% visual, 40% auditory, 20%

multisensory). **(d)** Schematic outlining the computations of a Bayesian ideal observer. Stimulus with a true

underlying rate s gives rise to noisy observations x_A and x_V, which are then integrated with each other

and with prior beliefs to form a multisensory posterior belief about the rate. This is combined with prior

beliefs about choice to produce a posterior belief about the correct choice, and further combined with reward

information to form expected action values Q_L, Q_R. The ideal observer selects the action $\hat{c}$ with maximum

expected value. Lightning bolts denote proposed sources of noise that can give rise to (red) or exacerbate

(grey) lapses, causing deviations from the ideal observer. **(e)** Posterior beliefs on an example trial assuming

flat priors. Solid black line denotes true rate, blue and green dotted lines denote noisy visual and auditory

observations, with corresponding unisensory posteriors shown in solid blue and green. Solid red denotes the

multisensory posterior, centered around the M.A.P. rate estimate in dotted red. Shaded fraction denotes the

probability of the correct choice being rightward, with μ denoting the category boundary. **(f)** Ideal observer

predictions for the psychometric curve, i.e. proportion of rightward choices for each rate. Inverse slopes of

the curves in each condition are reflective of the posterior widths on those conditions, assuming flat priors.

The value on the abscissa corresponding to the curve's midpoint indicates the category boundary (assuming

129   equal rewards).

Lapses cause deviations from ideal observer, and are reduced on multisensory trials

131   In practice, the shapes of empirically obtained psychometric curves do not perfectly match the

132   ideal observer (Fig. 2e, green) since they asymptote at values that are less than 1 or greater than

133   0. This is a well known phenomenon in psychophysics (Wichmann and Hill, 2001), requiring two

134   additional lapse parameters to precisely capture the asymptotes. To account for lapses, we fit a

135   four-parameter psychometric function to the subjects' choice data (Fig. 2a, Equation 1 in Methods)

136   with the Palamedes toolbox (Prins and Kingdom, 2018). γ and λ are the lower and upper asymptote

137   of the psychometric function, which parameterize lapses away from the left and the right reward

138   ports, respectively; ϕ is a sigmoidal function, in our case the cumulative normal distribution; x is

139   the event rate, i.e. the number of flashes or beeps presented during the one second stimulus period;

140   μ parameterizes the midpoint of the psychometric function and σ describes the inverse slope after

141   correcting for lapses.

142   How can we be sure that the asymptotes seen in the data truly reflect non-zero asymptotes

143   rather than fitting artifacts or insufficient data at the asymptotes? To test whether lapses were truly

144   necessary to explain the behavior, we fit the curves with and without lapses (Fig. 2b) and tested

145   whether the lapse parameters were warranted. The ideal observer without lapses was rejected in

146   15/17 rats by BIC, and in all rats by AIC. Fitting a fixed lapse rate was not sufficient to capture

147   the data, nor was fitting a lapse rate that was constrained to be less than 0.1 (Wichmann and Hill,

148   2001), and the data warranted fitting separate lapse rates to each condition ("free lapses" model

8

149 outperforms "fixed lapses", "restricted lapses" or "no lapses/ideal observer" in Fig. 2g).

150 Multisensory trials offer an additional, strong test of ideal observer predictions. In addition to

151 perfect performance on the easiest stimuli, the ideal observer model predicts the minimum possible

152 uncertainty achievable on multisensory trials through optimal integration (Ernst and Bulthoff, 2004;

153 Equation 2 in Methods). By definition, better-than-optimal performance is impossible. However,

154 studies in rats performing multisensory decision-making tasks suggest that in practice, animals'

155 performance appears to exceed optimal predictions (Raposo, Sheppard, et al., 2012; Nikbakht

156 et al., 2018) seeming, at first, to violate the ideal observer model. Moreover, in these datasets,

157 performance on the easiest stimuli was not perfect and asymptotes deviated from 0 and 1. As in

158 these previous studies, when we fit performance without lapses, multisensory performance was

159 significantly supra-optimal (p=0.0012, paired t-test), i.e. better than the ideal observer prediction

160 (Fig. 2c, data points are above the unity line). This was also true when lapse probabilities were

161 assumed to be fixed across conditions (p =0.0018) or when they were assumed to be less than 0.1

162 (p=0.0003). However, when we allowed lapses to vary freely across conditions, performance was

163 indistinguishable from optimal (Fig. 2d, data points are on the unity line). This reaffirms that proper

164 treatment of lapses is crucial for accurate estimation of perceptual parameters and offers a potential

165 explanation for previous reports of supra-optimality.

166 Using this improved fitting method, we replicated previous observations (Raposo, Sheppard, et

167 al., 2012; Raposo, Kaufman, and Churchland, 2014) showing that animals have improved sensitivity

168 (lower σ) on multisensory vs. unisensory trials (Fig. 2e, red curve is steeper than green/blue curves;

169 Fig. 2f, top). Interestingly, we observed that animals also had a lower lapse probability ($\lambda + \gamma$)

170 on multisensory trials (Fig. 2e, asymptotes for red curve are closer to 0 and 1; n=17 rats, 347537

171 trials). This was consistently observed across animals (Fig. 2f bottom, the probability of lapses on

172 multisensory trials was 0.06 on average, compared to 0.17 on visual, p=1.4e-4 and 0.21 on auditory,

173 p=1.5e-5).

174
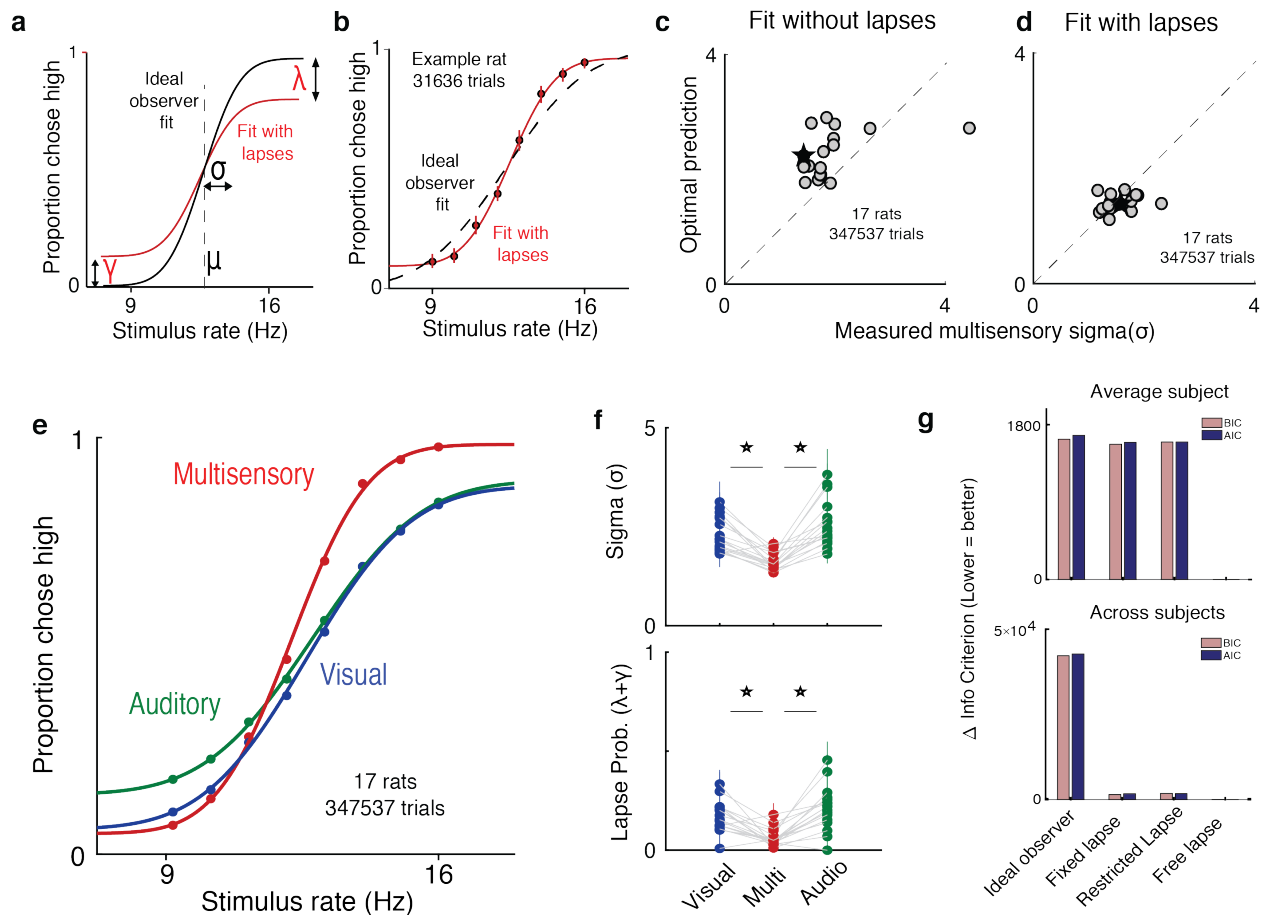
Figure 2   Deviations from ideal observer reflect lapses in judgment. **(a)** Schematic psychometric per-

176 formance of an ideal observer (black) vs. a model that includes lapses (green). The ideal observer fit

177 includes two parameters: intercept (μ) and slope (σ). The four-parameter cumulative normal distribution

178 individually fits μ, σ, and lapse probability for high rate (γ) and low rate choices (λ). Dotted line shows

the category boundary (12.5 Hz). **(b)** Subject data was fit with a an ideal observer model (black) and a four-parameter model (green). Dotted line shows the category boundary (12.5 Hz). **(c,d)** Ideal observer predictions vs. measured multisensory sigma for fits with and without lapses. **(c)** Multisensory integration seems supra-optimal when not accounting for lapses. **(d)** Optimal multisensory integration is restored when accounting for lapses. (n = 17 rats. Points represent individual rats; star represents pooled data across all rats. Data points that lie on the unity line represent cases in which the measured sigma was equal to the optimal prediction). **(e)** Rats' psychometric curves on auditory (green), visual (blue) and multisensory (red) trials. Bold lines represent data pooled across 17 rats; thin lines represent individual rats. **(f)** Fit values of sigma (top) and lapse parameters (bottom) on unisensory and multisensory conditions. Both parameters showed significant reduction on the multisensory conditions (paired t-test, $p<0.05$); n=17 rats (347537 trials). **(g)** Model comparison using BIC(pink) and AIC(blue) for the average subject (top) and across individual subjects (bottom). Lower scores indicate better fits. Both metrics favor a model where lapses are allowed to vary freely across conditions ("Free lapse") over one without lapses ("Ideal observer"), one with a fixed probability of lapses ("Fixed lapse") or where the lapses are restricted to being less than 0.1 ("Restricted lapse").

A novel model, uncertainty-guided exploration, explains lapses better than traditional models of inattention or motor-error

What could account for the reduction in lapse probability on multisensory trials? While adding extra parameters to the ideal observer model fit the behavioral data well and accounted for the reduction in inverse-slope on multisensory trials, this success doesn't provide an explanation for why lapses are present in the first place, nor why they differ between stimulus conditions.

11

200 To investigate this, we considered possible sources of noise that have traditionally been

201 invoked to explain lapses (Fig. 1d). We first hypothesized that lapses might be due to a fixed

202 amount of noise added once the decision has been made. These sources of noise could include

203 decision noise due to imprecision (Findling et al., 2018), motor errors (Wichmann and Hill, 2001)

204 or ϵ-greedy exploration. However, these sources should hinder decisions equally across conditions

205 (Supplementary Fig. 1b), which cannot explain our observation of condition-dependent lapse rates

206 (Fig. 2f).

207 A second explanation is that lapses arise due to inattention on a small fraction of trials.

208 Inattention would drive the animal to guess randomly, producing lapse rates whose sum should

209 reflect the probability of not attending (Fig. 3a, Methods). According to this explanation, the

210 lower lapse rate on multisensory trials reflects increased attention on those trials, perhaps due to

211 their increased bottom-up salience (i.e. two streams of stimuli instead of one). To test this, we

212 leveraged a multisensory condition that manipulates uncertainty without changing salience (Raposo,

213 Sheppard, et al., 2012). Specifically, we interleaved standard matched-rate multisensory trials with

214 "neutral" multisensory trials for which the rate of the auditory stimuli ranged between 9-16 Hz,

215 while the visual stimuli was always 12 Hz. This rate was so close to the category boundary (12.5

216 Hz) that it did not provide compelling evidence for one choice or the other (Fig. 3d, left), thus

217 reducing the information in the multisensory stimulus and increasing uncertainty. However, since

218 both "neutral" and "matched" conditions are multisensory, they should be equally salient, and since

219 they are interleaved, the animal would be unable to identify the condition without actually attending

220 to the stimulus. According to the inattention model, matched and neutral trials should have the same

221 rate of lapses, only differing in their σ (Supplementary Fig 1c).

222 Contrary to this prediction, we observed higher lapse rates on "neutral" trials, where the
223 uncertainty was high, than on "matched" trials, where the uncertainty was lower (Fig. 3d). The
224 dependence of lapses on uncertainty is reminiscent of the dependence of lapse on uncertainty
225 observed when comparing unisensory vs. multisensory trials (Fig. 2e,f; Supplementary Fig. 1e).

226 Having observed that traditional explanations of lapse fail to account for the behavioral
227 observations, we extended the ideal observer model to propose a novel explanation for lapses: an
228 uncertainty-guided form of exploration using a dynamic softmax decision rule (Fig. 3b). This form
229 of exploration is widely used in value-based decision making (Dayan and Daw, 2008) since it allows
230 the subject to "tune" the degree of exploration using a β parameter, also known as an "inverse
231 temperature". Lower values of β encourage more exploration, while higher values of β encourage
232 exploitation, with the limiting case of $\beta \to \infty$ reducing to the reward-maximizing ideal observer.
233 Modulating exploration with uncertainty is a well known heuristic called Thompson sampling, that
234 automatically balances exploration and exploitation (Supplementary Fig. 4a-b) allowing one to
235 efficiently maximize long-term expected reward (Gershman, 2018; Leike et al., 2016). This predicts
236 that conditions with higher uncertainty in expected reward (e.g. unisensory or neutral trials) should
237 encourage more exploration, giving rise to more frequent lapses (Supplementary Fig. 4c-d).

238 As a result, the uncertainty-guided exploration model predicts an increase not only in the σ
239 but also in lapses on neutral trials, just as we observed (Fig. 3c)- in fact it predicts that both these
240 parameters should match those on auditory trials. This model fit the data well (Fig. 3e, bottom).

13

241 The inattention model, by contrast, predicts that both conditions would have the same lapse rates

242 with the neutral condition simply having a greater σ. This model provided a worse fit to the data,

243 particularly missing the data at extreme stimulus values where lapses are most clearly apparent

244 (Fig. 3e, top). Model comparison using BIC and AIC both favored the exploration model over the

245 inattention model for average data (Fig. 3f top) as well as across individual subjects (Fig. 3f bottom,

246 Supplementary Fig. 3)
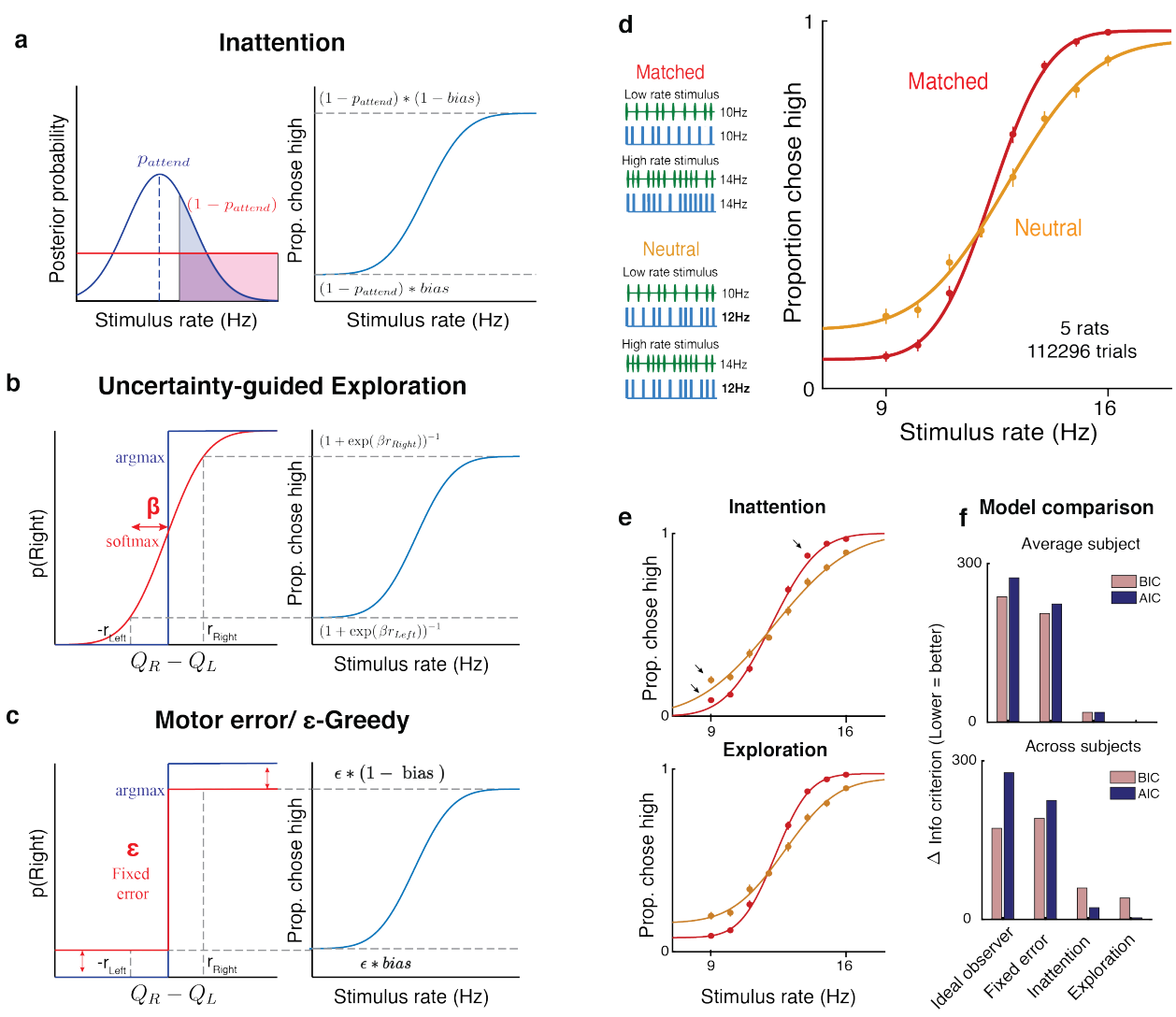
247

248 **Figure 3 Uncertainty-guided exploration offers a novel explanation for lapses.** Models of lapses in

249   decision-making: **(a)** Inattention model of lapses. Left panel: on a small fraction of trials given by $1 - p_{attend}$,

250   the observer does not attend to the stimulus (red curve), leading to equal posterior probabilities of high and

251   low rates (Shaded, clear regions of curve respectively) and guesses according to the probability $bias$, giving

252   rise to lapses (right panel). The sum of lapse rates then reflects $1 - p_{attend}$, while their ratio reflects the

253    $bias$. **(b)** Uncertainty-guided exploration model. Lapses can arise from exploratory decision rules such as

254   the "softmax" (red) rather than reward-maximization (blue). Since the difference in expected value from

255   right and left actions ($Q_R - Q_L$) is bounded by the maximum reward magnitudes r_R and r_L, even when

256   the stimulus is very easy, the maximum probability of choosing the higher value option is not 1, giving rise

257   to lapses. Lapse rates on either side are then proportional to the reward magnitude on that side, and to a

258   "temperature" parameter β that depends on the uncertainty in expected reward. **(c)** Motor error, or ϵ-greedy

259   model. Lapses can also arise from decision rules with a fixed proportion ϵ of random exploratory choices,

260   or due to motor errors ocurring on ϵ fraction of trials. The sum of lapses reflects ϵ while their ratio reflects

261   any $bias$ in exploration or motor errors. **(d)** Left: multisensory stimuli designed to distinguish between

262   attentional and non-attentional sources of lapses. Standard multisensory stimuli with matched visual and

263   auditory rates (top) and "neutral" stimuli where one modality has a rate very close to the category boundary

264   and is uninformative (bottom). Both stimuli are multisensory and designed to have equal bottom-up salience,

265   and can only be distinguished by attending to them and accumulating evidence. Right: rat performance

266   on interleaved matched (red) and neutral (orange) trials. **(e)** Since the matched and neutral conditions are

267   equally salient, they are expected to have equal probabilities of attending, predicting similar total lapse rates

268   in the inattention model (top, solid lines are model fits). Deviations from model fits are denoted with arrows.

269   The exploration model (bottom) provides a better fit, by allowing for different levels of exploration in the

two conditions. **(f)** Model comparison using BIC (pink) and AIC (blue) both favor the uncertainty-guided
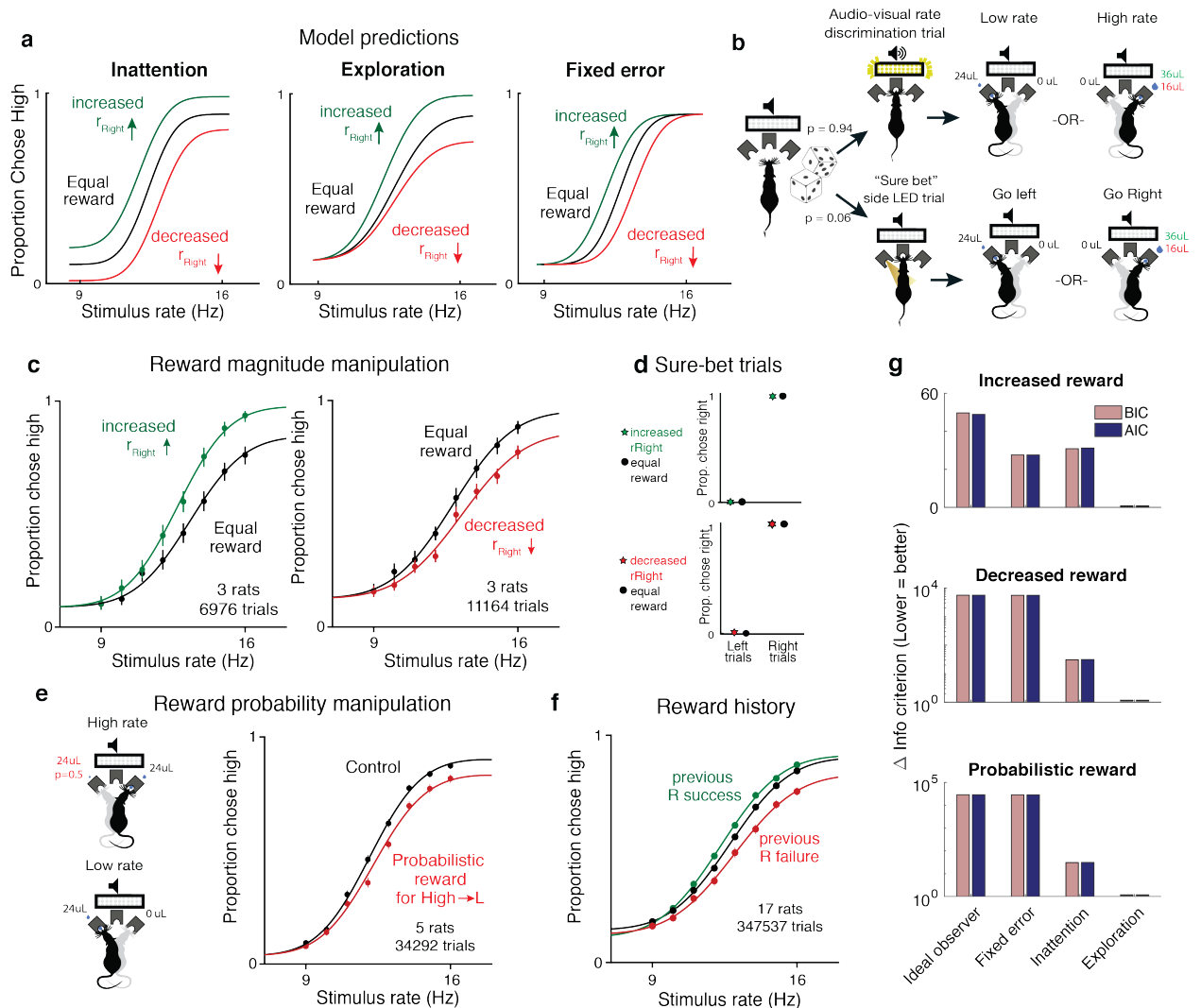
exploration model.

Reward manipulations confirm predictions of exploration model

One of the key features of the uncertainty-guided exploration model is that lapses are exploratory

choices made with full knowledge of the stimulus, and should depend only on the expected rewards

associated with that stimulus category (Supplementary Fig. 4). This is in stark contrast to the

inattention model and many other kinds of disengagement (Supplementary Fig. 2), in which lapses

are caused by the observer disregarding the stimulus, and hence lapses at the two extreme stimulus

levels are both influenced by a common underlying guessing process that depends on expected

rewards from both stimulus categories. This is also in contrast with fixed motor error or ϵ-greedy

models in which lapses are independent of expected reward (Fig. 3c).

Therefore, a unique prediction of the exploration model is that selectively manipulating

expected rewards associated with one of the stimulus categories should only affect lapses at one

extreme of the psychometric function, whereas inattention and other kinds of disengagement predict

that both lapses should be affected, and fixed error models predict that neither should be affected

(Fig. 4a, Supplementary Fig. 1,2).

To experimentally test these predictions, we tested rats on the rate discrimination task with

asymmetric rewards (Fig. 4b, top). Instead of rewarding high and low rate choices equally, we

increased the water amount on the reward port associated with high-rates (rightward choices) so it

was 1.5 times larger than before, without changing the reward on the the low-rate side (leftward

choices). In a second rat cohort we did the opposite: we devalued the choices associated with high-rate trials by decreasing the water amount on that side port so it was 1.5 times smaller than before, without changing the reward on the low-rate side.

Figure 4 **Reward manipulations match predictions of the exploration model**. **(a)** The inattention, exploration and fixed error models make different predictions for increases and decreases in the reward magnitude for rightward (high-rate) actions. The inattention model (left panel) predicts changes in lapses for both right and left choices, while the exploration model (center panel) predicts changes in lapses only for

298   high rate choices, and fixed motor error or ϵ greedy models (right) predict changes in neither lapse. Black

299   line, equal rewards on both sides; green, increased rightward reward; red, decreased rightward reward. **(b)**

300   Schematic of rate discrimination trials and interleaved "sure bet" (side LED) trials. The majority of the trials

301   (94%) were rate discrimination trials as described in Figure 1. On sure-bet trials, a pure tone was played

302   during a 0.2 second fixation period and one of the side ports was illuminated once the tone ended to indicate

303   that reward was available there. Rate discrimination and sure-bet trials were randomly interleaved, as were

304   left and right trials, and the rightward reward magnitude was either increased ($36\mu l$) or decreased ($16\mu l$) while

305   maintaining the leftward reward at $24\mu l$ **(c)** Rats' behavior on rate discrimination trials following reward

306   magnitude manipulations. High rate lapses decrease when water reward for high-rate choices is increased (left

307   panel; n=3 rats, 6976 trials), while high-rate lapses increase when reward on that side is decreased (right panel;

308   n=3 rats, 11164 trials). Solid curves are exploration model fits with a single parameter change accounting for

309   the manipulation. **(d)** Rats show nearly perfect performance on sure-bet trials, and are unaffected by reward

310   manipulations on these trials. **(e)** Reward probability manipulation. (Left) Schematic of probabilistic reward

311   trials, incorrect (leftward) choices on high rates were rewarded with a probability of 0.5, and all other rewards

312   were left unchanged. (Right) Rat behavior and exploration model fits showing a selective increase in high-rate

313   lapses (n=5 rats, 34292 trials). **(f)** Rat behavior on equal reward trials conditioned on successes (green) or

314   failures (red) on the right on the previous trials resembles effects of reward size manipulations. **(g)** Model

315   comparison showing that AIC and BIC both favor the exploration model on data from all 3 manipulations.

316   The animals' behavior on the asymmetric-reward task matched the predictions of the explo-

317   ration model. Increasing the reward size on choices associated with high-rates led to a decrease in

318   lapses for the highest rates and no changes in lapses for the lower rates (Fig. 4c, left; n=3 rats, 6976

18

319 trials). Decreasing the reward of choices associated with high-rates led to an increase in lapses for

320 the highest rates and no changes in lapses for the lower rates (Fig. 4c, right; n=3 rats, 11164 trials).

321 This shows that both increasing and decreasing the value of one of the actions has an asymmetric

322 effect on lapse probabilities that does not match the inattention model.

323   To confirm that the asymmetric changes in lapse rate that we observed were truly driven

324 by uncertainty, we examined performance on randomly interleaved "sure bet" trials on which the

325 uncertainty was very low (Fig. 4b, bottom). On these trials, a pure tone was played during the

326 fixation period, after which an LED at one of the side ports was clearly illuminated, indicating a

327 reward. Sure-bet trials comprised 6% of the total trials, and as with the rate discrimination trials,

328 left and right trials were interleaved. Owing to the low uncertainty, the model predicts that very

329 little exploration would be required in this condition, and that animals would very quickly reach

330 perfect performance on these trials. Importantly, our model predicts that performance on "sure-bet"

331 trials would be unaffected by imbalances in reward magnitude.

332   In keeping with this prediction, on sure-bet trials, performance was near perfect (right-

333 ward probabilities of 0.003 [0.001,0.01] and 0.989 [0.978,0.995] on go-left and go-right trials

334 respectively), and unaffected following reward manipulations (Fig. 4d: Rightward probabilities of

335 0.004 [0.001, 0.014] and 0.996 [0.986,0.999] on increased reward, 0.006 [0.003,0.012] and 0.99

336 [0.983,0.994] on decreased reward). This suggests that the effect of value on lapses is restricted

337 to uncertain situations that encourage subjects to explore, rather than exploit. Further, because

338 sure-bet trials were interleaved with more uncertain trials, their near-perfect performance indicates

339 that uncertainty can be estimated on the timescale of individual trials.

340   As an additional test of the model, we manipulated expected rewards by probabilistically

341 rewarding incorrect i.e. leftward choices on high rate trials with a probability of 0.5, while leaving all

342 other rewards unchanged (Fig. 4e left). The exploration model predicts that this should selectively

343 increase the value of leftward actions on high rate trials, increasing lapses on high rates. Indeed,

344 this is what we observed (Fig. 4e right, n=5 animals, 347537 trials), and the effect was strikingly

345 similar to the decreased reward experiment, even though the two manipulations affect high rate

346 action values through changes on opposite reward ports. Moreover, this suggests that lapses reflect

347 changes in action value caused by changing either reward magnitudes or reward probabilities, as

348 one would expect from the exploration model.

349   The subjective value of actions may naturally change with experience, even without the

350 explicit reward manipulations described above. Throughout training, the animal uses outcomes of

351 previous trials to learn and update the expected rewards from various actions, allowing it to learn

352 the rules of the task (i.e. that high rate trials are rewarded on the right and so on). If such learning

353 processes continue to persist in trained animals, either due to incomplete training (Law and Gold,

354 2009), uncertainty in feedback, forgetting over time (Gershman, 2015; Drugowitsch and Pouget,

355 2018), or perceived volatility (Yu and Cohen, 2009), then the outcomes of previous trials should

356 continue to affect subsequent trials even in trained animals, as has been observed in a number of

357 studies (Busse et al., 2011; Lak et al., 2018; Mendonca et al., 2018; Odoemene et al., 2018; Pinto

358 et al., 2018; Scott et al., 2015). The action value of rightward choices should increase following

359 a rightward success and decrease following a rightward failure, predicting the same asymmetric

360 changes in lapses as reward magnitude manipulations. As predicted, trials following rewarded and

361 unrewarded rightward choices showed decreased and increased high-rate lapses, respectively (Fig.

362 4g; same rats and trials as in Fig. 2e). Taken together, manipulations of value match the predictions

363 of the uncertainty-dependent exploration model.

Changes in lapses following prefrontal and striatal inactivations resemble value deficit

365 The results of the reward experiments suggest that disrupting areas that confer value to actions

366 should asymmetrically bias lapses, in contrast to disruptions of areas that encode sensory evidence,

367 which should lead to horizontal biases without affecting lapses, or motor disruptions that simply

368 make one of the actions harder to perform, which should affect both lapses (Supplementary Fig.

369 11a, top). Crucially, in the absence of lapses, all three of these disruptions would look identical,

370 producing a horizontal shift. This suggests that lapses could actually be informative about the

371 stage of involvement of brain regions. Two candidate areas that we sought out to test in our

372 multisensory task were secondary motor cortex (M2) and posterior striatum (pStr), both of which

373 receive convergent input from primary visual and auditory cortices (Supplementary Fig. 5, results of

374 simultaneous anterograde tracing from V1 and A1; also see Jiang and Kim, 2018; Barthas and Kwan,

375 2017). Previous studies have shown effects on lapses following inactivation of both these areas in

376 auditory tasks in rats (Erlich et al., 2015; Guo et al., 2018). These were interpreted as effects arising

377 from either leaky accumulation (Erlich et al., 2015), post-categorization biases (Piet et al., 2017) or

378 perceptual biases (Guo et al., 2018). These effects were very similar to the effects of manipulating

379 reward in our task, hinting that these effects may actually arise from biased action values. However,

380 since these deficits were only demonstrated during auditory decision-making, these studies did

381 not afford the opportunity to distinguish sensory modality-specific deficits from those that should

382 generalize across modalities (e.g., visual, multisensory, auditory) like value deficits.

383   To test whether pStr and M2 have a modality-independent role in perceptual decisions, we

384 suppressed activity of neurons in each of these areas using muscimol, a $GABA_A$ agonist, during

385 our multisensory rate discrimination task. We implanted bilateral cannulae in M2 (Supplementary

386 Fig. 6b; n = 5 rats; +2 mm AP 1.3 mm ML, 0.3 mm DV) and pStr (Supplementary Fig. 6a; n

387 = 6 rats; -3.2 mm AP, 5.4 mm ML, 4.1 mm DV) (Fig. 5a). On control days, rats were infused

388 unilaterally with saline, followed by unilateral muscimol infusion the next day (M2: 0.1-0.5 μg,

389 pStr 0.075-0.125 μg). We compared performance on the multisensory rate discrimination task for

390 muscimol days with preceding saline days. Inactivation of the side associated with low-rate choices

391 biased the animals to make more low-rate choices (Fig. 5b; left 6 panels: empty circles, inactivation

392 sessions; full circles, control sessions) and inactivation of the side associated with high-rates biased

393 them to make more high-rate choices (Fig. 5b, right 6 panels). The inactivations largely affected

394 lapses on the contralateral side, while sparing those on the ipsilateral side (Fig. 5c). These results

395 recapitulated previous findings, and were strikingly similar to the effects we observed following

396 reward manipulations (as seen in Fig. 4c, right panel). These effects were seen across areas (Fig. 5b,

397 top, M2; bottom, pStr) and modalities (Fig. 5b; green, auditory; blue, visual and red, multisensory),

398 suggesting that pStr and M2 are part of a modality-independent circuit for decision-making.

399   Fitting the data with the exploration model revealed that the effects on lapses could be captured

22

400 by scaling the contralateral action value by a single parameter across modalities (Fig. 5b, Joint fits

401 to control (solid lines) and inactivation trials (dotted lines) across modalities, differing only by a

402 single parameter), similar to the reward manipulation experiments. Animals that were inactivated

403 on the side associated with high rates showed increased low rate lapses (Fig. 5c, bottom right; data

404 points are above the unity line; n=9 rats), but lapses did not change for high rates (Fig. 5c, top right;

405 data points are on the unity line). This was consistent across areas and modalities (Fig. 5c; M2,

406 triangles; pStr, circles; blue, visual; green, auditory). Animals that were inactivated on the side

407 associated with low rates showed the opposite effect: increased lapses on high rate trials (Fig. 5c,

408 top left; n=10 rats) and no change in lapses for low rate trials (Fig. 5c bottom left). To confirm

409 that this effect was independent of the associated stimulus, some rats were trained on a reverse

410 contingency regimen in which high rates were rewarded on the left side. The effects were consistent

411 across both groups (Supplementary Fig. 7) and always resembled a devaluation of contralateral

412 actions (Supplementary Fig. 11).

413   To determine whether changes in decision-making simply reflected motor impairments that

414 drove a tendency to favor the ipsilateral side, we compared behavior on the sure-bet task described

415 previously (Fig. 4b, bottom). Performance was spared on these trials (Fig. 5d): rats made correct

416 rightward and leftward choices regardless of the side that was inactivated. This suggests that

417 behavioral effects were restricted to situations in which there was uncertainty about the correct

418 outcome. We also looked at multiple movement parameters such as wait time in the center port and

419 movement times to ipsilateral and contralateral reward ports. There were no significant effects on

420 movement parameters (Supplementary figure 10), suggesting that effects on decision outcome were

23

421   not due to a muscimol-induced motor impairment.

422   Finally, a model comparison revealed that a fixed contralateral value deficit captured the

423   inactivation effects much better than a fixed reduction in contralateral sensory evidence or a fixed

424   increase in contralateral motor effort, both for M2 (Fig. 5e top) and pStr (Fig. 5e bottom).

425   In uncertain conditions, this reduced contralateral value gives rise to more exploratory choices

426   and hence more lapses on one side (Fig. 5f top), but doesn't affect the other side, or sure-bet

427   trials on which the animals largely exploit. Together, this suggests that M2 and pStr have a

428   lateralized, modality-independent role in computing the expected value of actions based on incoming

429   multisensory information (Fig. 5f bottom).

430

24
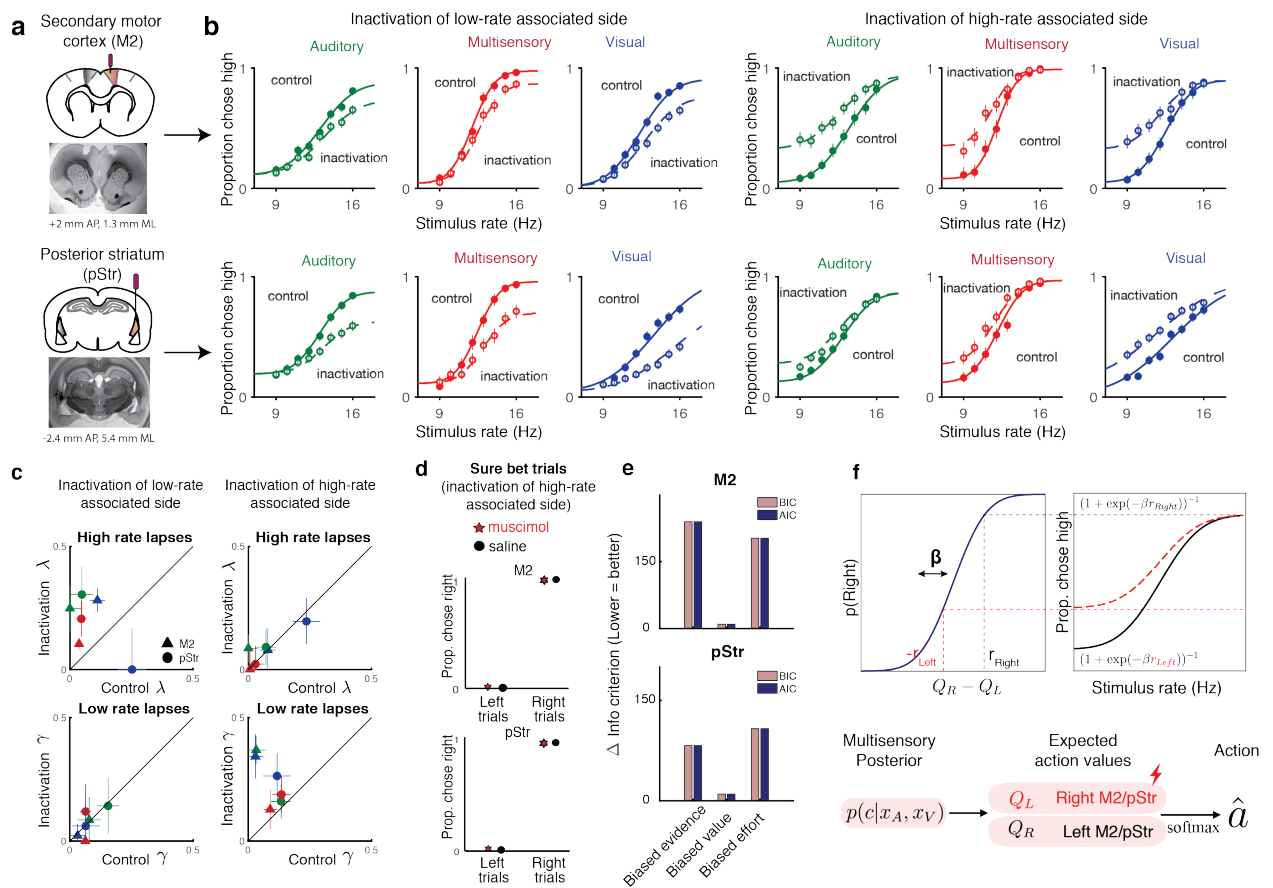
431 **Figure 5   Inactivation of secondary motor cortex and posterior striatum affects lapses, suggesting a**

432 **role in action value encoding**. **(a)** Schematic of cannulae implants in M2 (top) and pStr (bottom) and repre-

433 sentative coronal slices. For illustration purposes only, the schematic shows implants in the left hemisphere,

434 however, the inactivations shown in panel (b) were performed also on the right hemisphere. **(b)** Unilateral

435 inactivation of M2 (top) and pStr (bottom). Left 6 plots: inactivation of the side associated with low-rates

436 shows increased lapses for high rates on visual (blue), auditory (green) and multisensory (red) trials (M2: n=5

437 rats; 10329 control trials, full line; 6174 inactivation trials, dotted line; pStr: n=5 rats; 10419 control trials;

438 6079 inactivation trials). Right 6 plots: inactivation of the side associated with high-rates shows increased

439 lapses for low rates on visual, auditory and multisensory trials (M2: n=3 rats; 5678 control trials; 3816

440 inactivation trials; pStr: n=6 rats; 11333 control trials; 6838 inactivation trials). Solid lines are exploration

441 model fits,accounting for inactivation effects across all 3 modalities by scaling all contralateral values by a

442 single parameter. **(c)** Increased high rate lapses following unilateral inactivation of the side associated with

443 low-rates (top left); no change in low rate lapses (bottom left) and vice versa for inactivation of the side

444 associated with high-rates (top, bottom right). Control data on the abscissa is plotted against inactivation data

445 on the ordinate. Same animals as in **b**. Green, auditory trials; blue, visual trials. Abbreviations: posterior

446 striatum (pStr), secondary motor cortex (M2). **(d)** Sure bet trials are unaffected following inactivation. This

447 example shows that rats who were inactivated on the side associated with high rates make correct rightward

448 and leftward choices Top, M2; bottom, pStr. **(e)** Model comparison of three possible multisensory deficits -

449 reduction of contralateral evidence by a fixed amount (left), reduction of contralateral value by a fixed amount

450 (center) or an increased contralateral effort by a fixed amount (right). Both AIC and BIC suggest a value

451 deficit **(f)** Proposed computational role of M2 and Striatum. Lateralized encoding of left and right action

452 values by right and left M2/pStr explain the asymmetric effect of unilateral inactivations on lapses.

DISCUSSION

454 Perceptual decision-makers have long been known to display a small fraction of errors even on easy

455 trials. Until now, these "lapses" were largely regarded as a nuisance and lacked a comprehensive,

456 normative explanation. Here, we propose a novel explanation for lapses: that they reflect a strategic

457 balance between exploiting known rewarding options and exploring uncertain ones. Our model

458 makes strong predictions for lapses under diverse decision-making contexts, which we have tested

459 here. First, the model predicts more lapses on conditions with higher uncertainty, such as unisensory

460 (Fig. 2) or neutral (Fig. 3), compared to multisensory or sure-bet conditions. Second, the model

461 predicts that asymmetric reward manipulations should only affect lapses on one side, sparing

462 decisions to the other side and sure-bet trials (Fig. 4). Finally, the model predicts that lapses should

463 be affected by perturbations to brain regions that encode action value. Accordingly, we observed

464 that unilateral inactivations of secondary motor cortex and posterior striatum similarly affected

465 lapses on one side across auditory, visual and multisensory trials (Fig. 5). Taken together, our model

466 and experimental data argue strongly that far from being a nuisance, lapses are informative about

467 animals' subjective action values and reflect a trade-off between exploration and exploitation.

468   Considerations of value have provided many useful insights into aspects of behavior that

469 seem sub-optimal at first glance from the perspective of perceptual ideal observers. For instance,

470 many perceptual tasks are designed with accuracy in mind - defining an ideal observer as one

26

471 who maximizes accuracy, in line with classical signal detection theory. However, in practice, the

472 success or failure of different actions may be of unequal value to subjects, especially if reward or

473 punishment is delivered explicitly, as is often the case with non-human subjects. This may give rise

474 to biases that can only be explained by an observer that maximizes expected utility (Dayan and

475 Daw, 2008). Similarly, reward outcomes on a given trial can influence decisions about stimuli on

476 subsequent trials through reinforcement learning, giving rise to serial biases. These biases occur

477 even though the ideal observer should treat the evidence on successive trials as independent (Busse

478 et al., 2011; Lak et al., 2018). Finally, when subjects can control how long they sample the stimulus,

479 subjects maximizing reward rate may choose to make premature decisions, sacrificing accuracy for

480 speed (Bogacz et al., 2006).

481 Here, we take further inspiration from considerations of value to provide a novel account for

482 lapses. We leveraged a well known phenomenon in value-based decisions: uncertainty dependent

483 exploration. Until now, this phenomenon has not been considered a candidate explanation for

484 lapses in perceptual decisions. Our results argue that lapses are not simply accidental errors made

485 as a consequence of attentional "blinks" or motor "slips", but can reflect a deliberate, internal

486 source of behavioral variability that facilitates learning and information gathering under uncertain

487 or non-stationary environments.

488 Although exploration no longer yields the maximum utility on any given trial, it is critical

489 for dynamic environments, and those in which there is uncertainty about probability of reward or

490 stimulus-response contingency (e.g., during learning). By encouraging subjects to sample multiple

491 options, exploration can potentially improve the subject's knowledge of the rules of the task, helping

492 it to increase future payoff, thus maximizing expected utility over a long period of time.

493   Balancing exploration and exploitation is computationally challenging, and the mechanism

494 we propose here, uncertainty dependent exploration, is a well-known, elegant heuristic for achieving

495 this balance. Also known as Thompson sampling, this strategy has been shown to be asymptotically

496 optimal in partially observable environments (Leike et al., 2016) and can be naturally implemented

497 through a sampling scheme where the subject samples action values from a learnt distribution

498 and then maximizes with respect to the sample (Gershman, 2018). This strategy predicts that

499 conditions with higher uncertainty should have higher exploration, and consequently higher lapse

500 rates, explaining the pattern of lapse rates we observed on unisensory vs. multisensory trials as well

501 as on neutral vs. matched trials. A lower rate of lapses on multisensory trials has also been reported

502 on a visual-tactile task in rats (Nikbakht et al., 2018) and a vestibular integration task in humans

503 (Bertolini et al., 2015) and can potentially account for the apparent supra-optimal integration that

504 has been reported in a number of studies (Nikbakht et al., 2018; Hou et al., 2018; Raposo, Sheppard,

505 et al., 2012). A strong prediction of uncertainty guided exploration is that the animal should always

506 exploit on conditions with no uncertainty, as we observed on sure-bet trials (Fig. 4d, 5d).

507   The model also predicts that exploration, and consequently lapses, should decrease with

508 training as the animal becomes more certain of the rules and expected rewards, explaining training-

509 dependent effects reported in primates (Law and Gold, 2009). It can also potentially explain why

510 children have higher lapse rates - they have been shown to be more exploratory in their decisions

511 than adults (Lucas et al., 2014).

512 A unique prediction of the exploration model is that it predicts lapse rates will sometimes

513 change asymmetrically for left and right decisions. For instance, changing the value associated

514 with one of the decisions (eg. high rate) should only affect lapses associated with that decision -

515 predicting fewer lapses on high rates if the rightward reward is increased, and more lapses if it is

516 decreased, or if leftward decisions are probabilistically rewarded on high rates. These predictions

517 are borne out (Fig. 4c), and rightward successes or failures on the previous trial have a similar effect.

518 The model also suggests that the asymmetric effects on lapses seen during unilateral inactivations of

519 prefrontal and striatal regions (Fig. 5b) arises from a selective devaluation of contralateral actions.

520 This interpretation reconciles a number of studies that have found asymmetric effects of inactivating

521 these areas during perceptual decisions (Erlich et al., 2015; Zatka-Haas et al., 2019; Wang et al.,

522 2018; Guo et al., 2018) with their established roles in encoding action value (Barthas and Kwan,

523 2017; Lee et al., 2015) during value-based decisions.

524 An open question that remains is how the brain might tune the degree of exploration in

525 proportion to uncertainty. An intriguing candidate for this is dopamine, whose tonic levels have

526 been shown to modulate exploration in mice on a lever-press task (Beeler et al., 2010), and context-

527 dependent song variability in songbirds (Leblois, Wendel, and Perkel, 2010). Dopaminergic genes

528 have been shown to predict individual differences in uncertainty-guided exploration in humans

529 (Frank et al., 2009), and dopaminergic disorders such as Parkinson's disease have been shown to

530 disrupt the uncertainty-dependence of lapses across conditions (Bertolini et al., 2015). Patients with

ADHD, another disorder associated with dopaminergic dysfunction, have been shown to display both increased perceptual variability and increased task-irrelevant motor output, a measure that correlates with lapses (Mihali et al., 2018). A promising avenue for future studies is to leverage the informativeness of lapses and the precise control of uncertainty afforded by multisensory tasks, in conjunction with perturbations or recordings of dopaminergic circuitry, to further elucidate the connections between perceptual and value-based decision making systems.

METHODS

Behavior

Animal Subjects and Housing All animal procedures and experiments were in accordance with the National Institutes of Healths Guide for the Care and Use of Laboratory Animals and were approved by the Cold Spring Harbor Laboratory Animal Care and Use Committee. Experiments were conducted with 34 adult male and female Long Evans rats (250-350g, Taconic Farms) that were housed with free access to food and restricted access to water starting from the onset of behavioral training. Rats were housed on a reversed light-dark cycle; experiments were run during the dark part of the cycle. Rats were pair-housed during the whole training period.

Animal training and behavioral task Rats were trained following previously established methods (Raposo 2012, Sheppard 2013, Raposo 2014, Licata 2017). Briefly, rats were trained to wait in the center port for 1000 ms while stimuli were presented, and to associate stimuli with left/right reward ports. Stimuli for each trial consisted of a series of events: auditory clicks from a centrally

30

positioned speaker, full-field visual flashes, or both together. Stimulus events were separated by either long (100 ms) or short (50 ms) intervals. For the easiest trials, all inter-event intervals were identical, generating rates that were 9 events/s (all long intervals) or 16 events/s (all short intervals). More difficult trials included a mixture of long and short intervals, generating stimulus rates that were intermediate between the two extremes and therefore more difficult for the animal to judge. The stimulus began after a variable delay following when the rats snout broke the infrared beam in the center port. The length of this delay was selected from a truncated exponential distribution (λ = 30 ms, minimum = 10 ms, maximum = 200 ms) to generate an approximately flat hazard function. The total time of the stimulus was usually 1000 ms. Trials of all modalities and stimulus strengths were interleaved. For multisensory trials, the same number of auditory and visual events were presented (except for a subset of neutral trials). Auditory and visual stimulus event times were generated independently, as our previous work has demonstrated that rats make nearly identical decisions regardless of whether stimulus events are presented synchronously or independently (Raposo, Sheppard, et al., 2012). For most experiments, rats were rewarded with a drop of water for moving to the left reward port following low-rate trials and to the right reward port following high rate trials. For muscimol inactivation experiments half of the rats were rewarded according to the reverse contingency. Animals typically completed between 700 and 1,200 trials per day. Most experiments had 18 conditions (3 modalities 8 stimulus strengths), leading to 29-50 trials per condition per day.

To probe the effect of uncertainty on lapses, rats received catch trials consisting of multisensory neutral trials, where only the auditory modality provided evidence for a particular choice, whereas

571   the visual modality provided evidence that was so close to the category boundary (12 Hz) that it did

572   not support one choice or the other (Raposo, Sheppard, et al., 2012)

573   To probe the effect of value on lapses, we manipulated either reward magnitude or reward

574   probability associated with high rates, while keeping low rate trials unchanged. To increase or

575   decrease reward magnitude associated with high rates, the amount of water dispensed on the right

576   port was increased or decreased to 36ul or 16 ul respectively, while the reward on the left port

577   was maintained at 24 ul. To manipulate reward probability, we occasionally rewarded rats on the

578   (incorrect) left port on high rate trials with a probability of 0.5. The right port was still rewarded

579   with a probability of 1 on high rates, and reward probabilities on low rate trials were unchanged (1

580   on the left port, 0 on the right).

581   **Analysis of behavioral data.**

582   *Psychometric curves.* Descriptive four-parameter psychometric functions were fit to choice data us-

583   ing the Palamedes toolbox (Prins and Kingdom, 2018). Psychometric functions were parameterized

584   as:

$$\psi(x; \mu, \sigma, \gamma, \lambda) = \phi(x; \mu, \sigma)(1 - \lambda - \gamma) + \gamma \tag{1}$$

585   where γ and λ are the lower and upper asymptote of the psychometric function, which parameterize

586   the lapse rates on the left and to the right, respectively; ϕ is a cumulative normal function; x is the

587   event rate, i.e. the number of flashes or beeps presented during the one second stimulus period; μ

588   parameterizes the x-value at the midpoint of the psychometric function and σ describes the inverse

589   slope. 95% Confidence intervals on these parameters were generated via bootstrapping based on

590   1000 simulations.

Modeling

Ideal observer model

593   We can specify an ideal observer model for our task using Bayesian Decision Theory (Dayan and

594   Daw, 2008). This observer maintains probability distributions over previously experienced stimuli

595   and choices, computes the posterior probability of each action being correct given its observations

596   and picks the action that yields the highest expected reward.

597   Let the true category on any given trial be c_{true}, the true stimulus rate be s_{true} and the animal's

598   noisy visual and auditory observations of s_{true} be x_V and x_A, respectively. We assume that the two

599   sensory channels are corrupted by independent gaussian noise with standard deviation σ_A and σ_V,

600   respectively, giving rise to conditionally independent observations.

$$p(x_A|s_{true}) = \mathcal{N}(s_{true}, \sigma_A), \quad p(x_V|s_{true}) = \mathcal{N}(s_{true}, \sigma_V),$$

$$p(x_A, x_V|s_{true}) = p(x_A|s_{true})p(x_V|s_{true}) \tag{2}$$

601   The ideal observer can use this knowledge to compute the likelihood of seeing the current trial's

602   observations as a function of the hypothesized stimulus rate s. This likelihood $\mathcal{L}$ is a gaussian

33

function of s with a mean given by a weighted sum of the observations x_A and x_V,:

$$\mathcal{L}(s) = p(x_A, x_V|s) = p(x_A|s)p(x_V|s)$$

$$\propto \mathcal{N}(\mu_M, \sigma_M)$$

$$\mu_M = w_A x_A + w_V x_V \tag{3}$$

$$\sigma_M = (\sigma_A^{-2} + \sigma_V^{-2})^{-\frac{1}{2}}$$

$$w_A = \frac{\sigma_M^2}{\sigma_A^2}, \quad w_V = \frac{\sigma_M^2}{\sigma_V^2}$$

The likelihood of seeing the observations as a function of the hypothesized category c, is given by marginalizing over all possible hypothesized stimulus rates. Let the experimentally imposed category boundary be μ_0, such that stimulus rates are considered high when $s > \mu_0$ and low when $s < \mu_0$. Then,

$$\mathcal{L}(c = \text{High}) = p(x_A, x_V|c = \text{High})$$

$$= \int_s p(x_A, x_V, s|c = \text{High})ds$$

$$= \int_s p(x_A, x_V|s)p(s|c = \text{High})ds \quad \because x_a, x_V \perp c|s \tag{4}$$

$$= \int_{s > \mu_0} p(x_A, x_V|s)ds$$

$$\propto 1 - \Phi(\mu_0; \mu_M, \sigma_M)$$

where Φ is the cumulative normal function. Using Bayes' rule, the ideal observer can then compute the probability that the current trial was high or low rate given the observations, i.e. the posterior

34

610   probability.

$$p(c|x_A, x_V) = \frac{p(x_A, x_V|c)p(c)}{p(x_A, x_V)}$$

$$\implies p(c = \text{High}|x_A, x_V) \propto p_{High}(1 - \Phi(\mu_0; \mu_M, \sigma_M)) \tag{5}$$

$$\implies p(c = \text{Low}|x_A, x_V) \propto p_{Low}\Phi(\mu_0; \mu_M, \sigma_M)$$

611   where p_{High} and p_{Low} are the prior probabilities of high and low rates respectively. The expected

612   value $Q(a)$ of choosing right or left actions (also known as the action values) is obtained by

613   marginalizing the learnt value of state-action pairs $q(c, a)$ over the unobserved state c.

$$Q(a = R) = p(\text{High}|x_A, x_V)q(High, R) + p(\text{Low}|x_A, x_V)q(Low, R)$$

$$\tag{6}$$

$$Q(a = L) = p(\text{High}|x_A, x_V)q(High, L) + p(\text{Low}|x_A, x_V)q(Low, L)$$

614   Under the standard contingency, high rates are rewarded on the right and low rates on the left,

615   so for a trained observer that has fully learnt the contingency, $q(High, R) \to r_R, q(High, L) \to$

616   $0, q(Low, R) \to 0, q(Low, L) \to r_L$, with rR and rL being reward magnitudes for rightward and

617   leftward actions. This simplifies the action values to:

$$Q(R) = p(\text{High}|x_A, x_V)r_R \propto p_{High}(1 - \Phi(\mu_0; \mu_M, \sigma_M))r_R$$

$$\tag{7}$$

$$Q(L) = p(\text{Low}|x_A, x_V)r_L \propto p_{Low}\Phi(\mu_0; \mu_M, \sigma_M)r_L$$

35

618   The max-reward decision rule involves picking the action $\hat{a}$ with the highest expected reward:

$$\hat{a} = \text{argmax} Q(a)$$

$$\text{i.e. } \hat{a} = R \iff Q(R) > Q(L)$$

$$\iff p_{High}(1 - \Phi(\mu_0; \mu_M, \sigma_M))r_R > p_{Low}\Phi(\mu_0; \mu_M, \sigma_M))r_L \tag{8}$$

$$\iff \Phi(\mu_M; \mu_0, \sigma_M)) > \frac{1}{1 + \frac{p_{High}r_R}{p_{Low}r_L}}$$

$$\iff w_A x_A + w_V x_V > \Phi^{-1}(\frac{1}{1 + \frac{p_{High}r_R}{p_{Low}r_L}}; \mu_0, (\sigma_A^{-2} + \sigma_V^{-2})^{-\frac{1}{2}})$$

619   In the special case of equal rewards and uniform stimulus and category priors, this reduces to

620   choosing right when the weighted sum of observations is to the right of the true category boundary,

621   i.e. $w_A x_A + w_V x_V > \mu_0$. Note that this is a deterministic decision rule for any given observations

622   x_A and x_V, however, since these are noisy and gaussian distributed around the true stimulus rate

623   s_{true}, the likelihood of making a rightward decision is given by the cumulative gaussian function Φ:

624

$$\text{For} \quad p_{High} = p_{Low}, r_R = r_L$$

$$p(\hat{a} = R|s) = p(w_A x_A + w_V x_V > \mu_0|s)$$

$$= \Phi(s_{true}; \mu_0, \sigma) \tag{9}$$

$$\sigma = \begin{cases} \sigma_A \text{ on auditory trials} \\ \sigma_V \text{ on visual trials} \\ (\sigma_A^{-2} + \sigma_V^{-2})^{\frac{1}{2}} \text{ on multisensory trials} \end{cases}$$

625

36

626 We can measure this probability empirically through the psychometric curve. Fitting it with a two

627 parameter cumulative gaussian function yields μ and σ which can be compared to ideal observer

628 predictions. The σ parameter is then taken to reflect sensory noise; and with the assumption of uni-

629 form priors and equal rewards, the μ parameter is taken to reflect the subjective category boundary.

630 Although μ should equal μ_0 for the ideal observer, in practice it is treated as a free parameter, and

631 deviations of μ from μ_0 could reflect any of three possible suboptimalities: 1) a subjective category

632 boundary mismatched to the true one, 2) mismatched priors, or 3) unequal subjective rewards of the

633 two actions.

634

Inattention model

636 The traditional model for lapse rates assumes that on a fixed proportion of trials, the animal fails to

637 pay attention to the stimulus, guessing randomly between the two actions. We can incorporate this

638 suboptimality into the ideal observer above as follows: Let the probability of attending be p_{attend}.

639 Then, on $1 - p_{attend}$ fraction of trials, the animal does not attend to the stimulus (i.e. receives

640 no evidence), effectively making $\sigma_{sensory} \to \infty$ and giving rise to a posterior that is equal to the

641 prior. On these trials, the animal may choose to maximize this prior (always picking the option

642 that's more likely a-priori, guessing with 50-50 probability if both options are equally likely), or

643 probability-match the prior (guessing in proportion to its prior). Let us call this guessing probability

644 p_{bias}. Then, the probability of a rightward decision is given by marginalizing over the attentional

645 state:

37

646

$$p(\hat{a} = R|s) = p(\hat{a} = R|s, \text{attend})p(\text{attend}) + p(\hat{a} = R|s, \sim \text{attend})p(\sim \text{attend})$$

$$= p(\hat{a} = R|s)p_{attend} + p_{bias}(1 - p_{attend}) \tag{10}$$

647 Comparing this with the standard 4-parameter sigmoid used in psychometric fitting, we obtain

$$p(\hat{a} = R|s_{true}) = \gamma + (1 - \gamma - \lambda)\Phi(s_{true}; \mu_0, \sigma)$$

$$\implies \gamma + \lambda = p_{attend}, \quad \frac{\gamma}{\gamma + \lambda} = p_{bias} \tag{11}$$

648 where γ and λ are the lower and upper asymptotes respectively, collectively known as "lapses".

649 In this model, the sum of the two lapses depends on the probability of attending, which could be

650 modulated in a bottom up fashion by the salience of the stimulus; their ratio depends on the guessing

651 probability, which in turn depends on the observer's priors and subjective rewards.

652

Motor error/ϵ greedy model

654 Lapses can also occur if the observer doesn't always pick the reward-maximizing or "exploit"

655 decision. This might occur due to random errors in motor execution on a small fraction of trials

656 given by ϵ, or it might reflect a deliberate propensity to occasionally make random "exploratory"

657 choices to gather information about rules and rewards. This is known as an ϵ-greedy decision rule,

658 where the observer chooses randomly (or according to p_{bias}) on ϵ fraction of trials. Both these

659 models yield predictions similar to those of the inattention model:

$$p(\hat{a} = R|s) = p(\hat{a} = R|s)(1 - \epsilon) + \epsilon p_{bias}$$

$$\implies \gamma + \lambda = \epsilon, \quad \frac{\gamma}{\gamma + \lambda} = p_{bias}$$

(12)

660

Softmax exploration model

662 A more sophisticated form of exploration is the "softmax" decision rule, which explores options in

663 proportion to their expected rewards, allowing for a balance between exploration and exploitation

664 through the tuning of a parameter β known as inverse temperature. In particular, in conditions of

665 greater uncertainty about rules or rewards, it is advantageous to be more exploratory and have a

666 lower β. This strategy is known as Thompson sampling, and can be achieved by sampling from a

667 belief distribution over expected rewards and maximizing with respect to the sample, reducing to a

668 softmax rule whose β depends on the total uncertainty in expected reward (Gershman, 2018).

$$p(\hat{a} = R|Q(a)) = \frac{\exp \beta Q(R)}{\exp \beta Q(L) + \exp \beta Q(R)}$$

$$= \frac{1}{1 + \exp(-\beta(Q(R) - Q(L)))}$$

(13)

669 The proportion of rightward choices conditioned on the true stimulus rate is then obtained

670 by marginalizing over the latent action values $Q(a)$, using the fact that the choice depends on s

671 only through its effect on $Q(a)$, where ρ is the animal's posterior belief in a high rate stimulus,

672 i.e. $\rho = p(c = High|x_A, x_V)$. ρ is often referred to as the *belief state* in reinforcement learning

673 problems involving partial observability such as our task.

$$
\begin{aligned}
p(\hat{a} = R|s) &= \int_{Q(a)} p(\hat{a} = R, Q(a)|s)dQ \\
&= \int_{Q(a)} p(\hat{a} = R|Q(a))p(Q(a)|s)dQ \quad \because \hat{a} \perp s|Q(a) \\
&= \int_{\rho} \frac{1}{1 + \exp -\beta(\rho(r_R + r_L) - r_L)} \frac{\mathcal{N}(\Phi^{-1}(1 - \rho, 0, \sigma_{post}), \mu_0 - s, \sigma_{post})}{\mathcal{N}(\Phi^{-1}(1 - \rho, 0, \sigma_{post}), 0, \sigma_{post}))} d\rho
\end{aligned}
\tag{14}
$$

674 Since lapses are the asymptotic probabilities of the lesser rewarding action at extremely easy

675 stimulus rates, we can derive them from this expression by setting $\rho \to 1$ or $\rho \to 0$. This yields

$$
\gamma = \frac{1}{1 + \exp(\beta r_L)}, \quad \lambda = \frac{1}{1 + \exp(\beta r_R)}
\tag{15}
$$

676 Critically, in this model, the upper and lower lapses are dissociable, depending only on the

677 rightward or leftward rewards, respectively. Such a softmax decision rule has been used to account

678 for suboptimalities in value based decisions (Dayan and Daw, 2008), however it has not been

679 used to account for lapses in perceptual decisions. Other suboptimal decision rules described

680 in perceptual decisions, such as generalized probability matching or posterior sampling (Acerbi,

681 Vijayakumar, and Wolpert, 2014; Drugowitsch, Wyart, et al., 2016; Ortega and Braun, 2013) amount

682 to a softmax on log-posteriors or log-expected values, rather than on expected values, and do not

683 produce lapses since in these decision rules, when the posterior probability goes to 1, so does the

684 decision probability.

685

Model fitting

Model fits were obtained from custom maximum likelihood fitting code using MATLAB's fmincon, by maximizing the marginal likelihood of rightward choices given the stimulus on each trial as computed from each model. Confidence intervals for fit parameters were generated using the hessian obtained from fmincon. Fits to multiple conditions were performed jointly, taking into account any linear or nonlinear (eg. optimality) constraints on parameters across conditions. Model comparisons were done using AIC and BIC.

Surgical procedures

All rats subject to surgery were anesthetized with 1%-3% isoflurane. Isoflurane anesthesia was maintained by monitoring respiration, heart rate, oxygen and CO_2 levels, as well as foot pinch responses throughout the surgical procedure. Ophthalmic ointment was applied to keep the eyes moistened throughout surgery. After scalp shaving, the skin was cleaned with 70% ethanol and 5% betadine solution. Lidocaine solution was injected below the scalp to provide local analgesia prior to performing scalp incisions. Meloxicam (5mg/ml) was administered subcutaneously (2mg/kg) for analgesia at the beginning of the surgery, and daily 2-3 days post-surgery. The animals were allowed at least 7 days to recover before behavioral training.

Viral injections- 2 rats, 15 weeks of age, were anesthetized and placed in a stereotaxic apparatus (Kopf Instruments). Small craniotomies were made in the center of primary visual cortex (V1; 6.9mm posterior to Bregma, 4.2mm to the right of midline) and primary auditory cortex (A1; 4.7mm posterior to Bregma, 7mm to the right of midline). Small durotomies were performed

at each craniotomy and virus was pressure injected at depths of 600, 800, and 1000 μm below the pia (150 nL/depth). Virus injections were performed using Drummond Nanoject III, which enables automated delivery of small volumes of virus. To minimize virus spread, the Nanoject was programmed to inject slowly: fifteen 10 nL boluses, 30 seconds apart. Each bolus was delivered at 10 nL/sec. 2-3 minutes were allowed following injection at each depth to allow for diffusion of virus. The AAV2.CB7.CI.EGFP.WPRE.RBG construct was injected in V1, and the AAV2.CAG.tdTomato.WPRE.SV40 construct was injected in A1. Viruses were obtained from the University of Pennsylvania vector core.

Cannulae implants Rats were anesthetized and placed in the stereotax as described above. After incision and skull cleaning, 2 skull screws were implanted to add more surface area for the dental cement. For striatal implants, two craniotomies were made, one each side of the skull (3.2mm posterior to Bregma; 5.4mm to the right and left of midline). Durotomies were performed and a guide cannula (22 gauge, 8.5 mm long; PlasticsOne) was placed in the brain, 4.1mm below the pia at each craniotomy. For secondary motor cortex implants, one large craniotomy spanning the right and left M2 was performed (~5mm x ~2mm in size centered around 2mm anterior to Bregma and 3.1mm to the right and left of midline). A durotomy was performed and a double guide cannula (22 gauge, 4mm long; PlasticsOne) was placed in the brain, 300um below the pia. The exposed brain was covered with sterile Vaseline and cannulae were anchored to the skull with dental acrylic (Relyx). Single or double dummy cannulae protruding 0.7 mm below the guide cannulae were inserted.

Inactivation with muscimol

Rats were lightly anesthetized with isoflurane. Muscimol was unilaterally infused into pStr or M2 with a final concentration of 0.075-0.125 μg and 0.1-0.5 μg, respectively. A single/double-internal cannula (PlasticsOne), connected to a 2 μl syringe (Hamilton microliter syringe, 7000 series), was inserted into each previously implanted guide cannula. Internal cannulae protruded 0.5mm below the guide. Muscimol was delivered using an infusion pump (Harvard PHD 22/2000) at a rate of 0.1 ul/minute. Internal cannulae were kept in the brain for 3 additional minutes to allow for diffusion of muscimol. Rats were removed from anesthesia and returned to cages for 15 minutes before beginning behavioral sessions. The same procedure was used in control sessions, where muscimol was replaced with sterile saline.

Histology

At the conclusion of inactivation experiments, animals were deeply anesthetized with Euthasol (pentobarbital and phenytoin). Animals were perfused transcardially with 4% paraformaldehyde. Brains were extracted and post-fixed in 4% paraformaldehyde for 24-48 hours. After post-fixing, 50-100 um coronal sections were cut on a vibratome (Leica) and imaged.

References

Acerbi, Luigi, Sethu Vijayakumar, and Daniel M Wolpert (2014). "On the origins of suboptimality in human probabilistic inference". In: *PLoS computational biology* 10.6, e1003661.

744 Barthas, Florent and Alex C Kwan (2017). "Secondary motor cortex- where sensory meets motor in

745 the rodent frontal cortex". In: *Trends in neurosciences* 40.3, pp. 181–193.

746 Bays, Paul M, Raquel FG Catalao, and Masud Husain (2009). "The precision of visual working

747 memory is set by allocation of a shared resource". In: *Journal of vision* 9.10, pp. 7–7.

748 Beeler, Jeff A et al. (2010). "Tonic dopamine modulates exploitation of reward learning". In:

749 *Frontiers in behavioral neuroscience* 4, p. 170.

750 Bertolini, Giovanni et al. (2015). "Impaired tilt perception in Parkinsons disease- a central vestibular

751 integration failure". In: *PloS one* 10.4, e0124253.

752 Bogacz, Rafal et al. (2006). "The physics of optimal decision making: a formal analysis of models of

753 performance in two-alternative forced-choice tasks." In: *Psychological review* 113.4, p. 700.

754 Busse, Laura et al. (2011). "The detection of visual contrast in the behaving mouse". In: *Journal of

755 Neuroscience* 31.31, pp. 11351–11361.

756 Carandini, Matteo and Anne K Churchland (2013). "Probing perceptual decisions in rodents". In:

757 *Nature neuroscience* 16.7, p. 824.

758 Dayan, Peter and Nathaniel D Daw (2008). "Decision theory, reinforcement learning, and the brain".

759 In: *Cognitive, Affective, & Behavioral Neuroscience* 8.4, pp. 429–453.

760 Drugowitsch, Jan and Alexandre Pouget (2018). "Learning optimal decisions with confidence". In:

761 *bioRxiv*, p. 244269.

762 Drugowitsch, Jan, Valentin Wyart, et al. (2016). "Computational precision of mental inference as

763 critical source of human choice suboptimality". In: *Neuron* 92.6, pp. 1398–1411.

Erlich, Jeffrey C et al. (2015). "Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat". In: *Elife* 4, e05457.

Ernst, Marc O and Heinrich H Bulthoff (2004). "Merging the senses into a robust percept". In: *Trends in cognitive sciences* 8.4, pp. 162–169.

Findling, Charles et al. (2018). "Computational noise in reward-guided learning drives behavioral variability in volatile environments". In: *bioRxiv*, p. 439885.

Frank, Michael J et al. (2009). "Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation". In: *Nature neuroscience* 12.8, p. 1062.

Garrido, Marta I, Raymond J Dolan, and Maneesh Sahani (2011). "Surprise leads to noisier perceptual decisions". In: *i-Perception* 2.2, pp. 112–120.

Gershman, Samuel J (2015). "A unifying probabilistic view of associative learning". In: *PLoS computational biology* 11.11, e1004567.

— (2018). "Deconstructing the human algorithms for exploration". In: *Cognition* 173, pp. 34–42.

Gold, Joshua I and Long Ding (2013). "How mechanisms of perceptual decision-making affect the psychometric function". In: *Progress in neurobiology* 103, pp. 98–114.

Green, David M, John A Swets, et al. (1966). *Signal detection theory and psychophysics*. Vol. 1. Wiley New York.

Guo, Lan et al. (2018). "Stable representation of sounds in the posterior striatum during flexible auditory decisions". In: *Nature communications* 9.1, p. 1534.

Hou, Han et al. (2018). "Neural correlates of optimal multisensory decision making". In: *bioRxiv*, p. 480178.

Jiang, Haiyan and Hyoung F Kim (2018). "Anatomical inputs from the sensory and value structures to the tail of the rat striatum". In: *Frontiers in neuroanatomy* 12.

Lak, Armin et al. (2018). "Dopaminergic and frontal signals for decisions guided by sensory evidence and reward value". In: *bioRxiv*, p. 411413.

Law, Chi-Tat and Joshua I Gold (2009). "Reinforcement learning can account for associative and perceptual learning on a visual-decision task". In: *Nature neuroscience* 12.5, p. 655.

Leblois, Arthur, Benjamin J Wendel, and David J Perkel (2010). "Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors". In: *Journal of Neuroscience* 30.16, pp. 5730–5743.

Lee, A Moses et al. (2015). "Between the primate and reptilian brain- rodent models demonstrate the role of corticostriatal circuits in decision making". In: *Neuroscience* 296, pp. 66–74.

Leike, Jan et al. (2016). "Thompson sampling is asymptotically optimal in general environments". In: *arXiv preprint arXiv:1602.07905*.

Licata, Angela M et al. (2017). "Posterior parietal cortex guides visual decisions in rats". In: *Journal of Neuroscience* 37.19, pp. 4954–4966.

Lucas, Christopher G et al. (2014). "When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships". In: *Cognition* 131.2, pp. 284–299.

Mendonca, Andre G et al. (2018). "The impact of learning on perceptual decisions and its implication for speed-accuracy tradeoffs". In: *bioRxiv*, p. 501858.

Mihali, Andra et al. (2018). "A Low-Level Perceptual Correlate of Behavioral and Clinical Deficits in ADHD". In: pp. 1–23.

Nikbakht, Nader et al. (2018). "Supralinear and supramodal integration of visual and tactile signals in rats: psychophysics and neuronal mechanisms". In: *Neuron* 97.3, pp. 626–639.

Odoemene, Onyekachi et al. (2018). "Visual evidence accumulation guides decision-making in unrestrained mice". In: *Journal of Neuroscience* 38.47, pp. 10143–10155.

Ortega, Pedro A and Daniel A Braun (2013). "Thermodynamics as a theory of decision-making with information-processing costs". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 469.2153, p. 20120683.

Piet, Alex T et al. (2017). "Rat prefrontal cortex inactivations during decision making are explained by bistable attractor dynamics". In: *Neural computation* 29.11, pp. 2861–2886.

Pinto, Lucas et al. (2018). "An accumulation-of-evidence task using visual pulses for mice navigating in virtual reality". In: *Frontiers in behavioral neuroscience* 12, p. 36.

Prins, Nicolaas and Frederick AA Kingdom (2018). "Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox". In: *Frontiers in psychology* 9.

Raposo, David, Matthew T Kaufman, and Anne K Churchland (2014). "A category-free neural population supports evolving demands during decision-making". In: *Nature neuroscience* 17.12, p. 1784.

Raposo, David, John P Sheppard, et al. (2012). "Multisensory decision-making in rats and humans". In: *Journal of neuroscience* 32.11, pp. 3726–3735.

Roy, Nicholas G et al. (2018). "Efficient inference for time-varying behavior during learning". In: *Advances in Neural Information Processing Systems*, pp. 5700–5710.

Scott, Benjamin B et al. (2015). "Sources of noise during accumulation of evidence in unrestrained and voluntarily head-restrained rats". In: *Elife* 4, e11308.

Shen, Shan and Wei Ji Ma (2019). "Variable precision in visual perception." In: *Psychological review* 126.1, p. 89.

Sheppard, John P, David Raposo, and Anne K Churchland (2013). "Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans". In: *Journal of vision* 13.6, pp. 4–4.

Wang, Lupeng et al. (2018). "Activation of striatal neurons causes a perceptual decision bias during visual change detection in mice". In: *Neuron* 97.6, pp. 1369–1381.

Wichmann, Felix A and N Jeremy Hill (2001). "The psychometric function: I. Fitting, sampling, and goodness of fit". In: *Perception & psychophysics* 63.8, pp. 1293–1313.

Witton, Caroline, Joel B Talcott, and G Bruce Henning (2017). "Psychophysical measurements in children: challenges, pitfalls, and considerations". In: *PeerJ* 5, e3231.

Yartsev, Michael M et al. (2018). "Causal contribution and dynamical encoding in the striatum during evidence accumulation." In: *Elife* 7.

Yu, Angela J and Jonathan D Cohen (2009). "Sequential effects: superstition or rational behavior?" In: *Advances in neural information processing systems*, pp. 1873–1880.

Zatka-Haas, Peter et al. (2019). "Distinct contributions of mouse cortical areas to visual discrimination". In: *bioRxiv*, p. 501627.

847 Zhou, Baohua et al. (2018). "Chance, long tails, and inference in a non-Gaussian, Bayesian theory

848   of vocal learning in songbirds". In: *Proceedings of the National Academy of Sciences* 115.36,

849   E8538–E8546.

858 **Competing Interests** The authors declare that they have no competing financial interests.

859 **Correspondence** Correspondence and requests for materials should be addressed to Anne K. Church-

860 land (email: churchland@cshl.edu).