Title
Lapses in perceptual decisions reflect exploration

Authors
Pisupati, Sashank
Chartarifsky-Lynn, Lital
Khanal, Anup
et al.

Publication Date
2021

DOI
10.7554/elife.55490

Peer reviewed

Lapses in perceptual decisions reflect exploration

Sashank Pisupati*[1,2], Lital Chartarifsky-Lynn*[1,2], Anup Khanal[1] & Anne K. Churchland[3]

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

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

[3]*University of California, Los Angeles, Los Angeles, California*

ABSTRACT

Perceptual decision-makers often display a constant rate of errors independent of evidence strength. These "lapses" are treated as a nuisance arising from noise tangential to the decision, e.g. inattention or motor errors. Here, we use a multisensory decision task in rats to demonstrate that these explanations cannot account for lapses' stimulus dependence. We propose a novel explanation: lapses reflect a strategic trade-off between exploiting known rewarding actions and exploring uncertain ones. We tested this model's predictions by selectively manipulating one action's reward magnitude or probability. As uniquely predicted by this model, changes were restricted to lapses associated with that action. Finally, we show that lapses are a powerful tool for assigning decision-related computations to neural structures based on disruption experiments (here, posterior striatum and secondary motor cortex). These results suggest that lapses reflect an integral component of decision-making and are informative about action values in normal and disrupted brain states.

INTRODUCTION

Perceptual decisions 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).

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; Busse et al., 2011; Gold and Ding, 2013; Carandini and Churchland, 2013). 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 tangential 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

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 learning in non-human primates (Law and Gold, 2009; Cloherty et al., 2019). 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; Flesch et al., 2018), 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 in healthy adult humans (Roach, Edwards, and Hogben, 2004; Witton, Talcott, and Henning, 2017; Manning et al., 2018). 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 decision-making task in rodents to reveal the inadequacy of traditional models. We challenge a key assumption of perceptual decision-making theories, i.e. subjects' perfect knowledge of expected rewards (Dayan and Daw, 2008), to uncover a novel explanation for lapses: uncertainty-guided exploration, a well known strategy for balancing exploration and exploitation in value-based decisions. We test predictions of the exploration model for perceptual decisions by manipulating the magnitude and probability of 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 contralateral

to 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, reflecting a core component of decision-making.

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) 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 of 12.5 Hz (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/sec; Fig. 1c, bottom) interleaved with visual-only or auditory-only trials. This allowed us to separately estimate the sensory noise in the animal's 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 noise in the animal's estimate: the higher the

standard deviation (σ) of sensory noise, the more uncertain the animal's 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 (Fig. 1e, visual [blue] and auditory [green] σ values are larger than multisensory σ [red]), driving a steeper psychometric curve (Fig. 1f, red curve is steeper than green and blue curves). Since this model does not take lapses into account, it would predict perfect performance on the easiest stimuli on all conditions, and thus all curves should asymptote at 0 and 1 (Fig 1f).

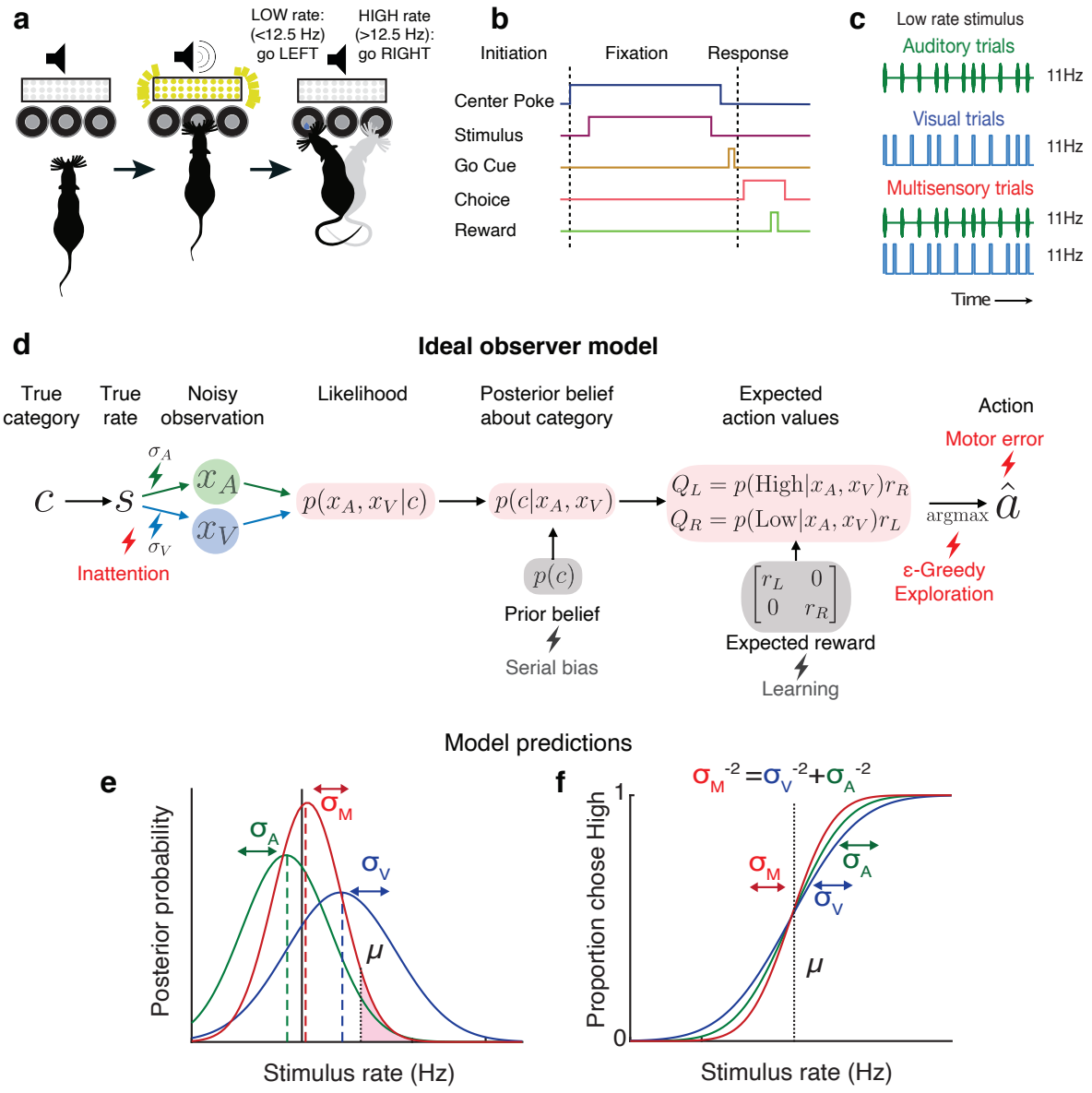


Figure 1 Testing ideal observer predictions in perceptual decision-making. (a) Schematic drawing of rate discrimination task. Rats initiate trials by poking into a center port. Trials consist of visual stimuli presented via a panel of diffused LEDs, auditory stimuli presented via a centrally positioned speaker or multisensory stimuli presented from both. Rats are rewarded with a 24 μl drop of water for reporting high rate stimuli (greater than 12.5 Hz) with rightward choices 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 belonging to a true category c, 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 category, and further combined with reward information to form expected action values Q_L, Q_R. The ideal observer selects the action $\hat{a}$ 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 maximum a posteriori 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 high rate 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 subjective category boundary, assuming equal rewards and flat priors.

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

In practice, the shapes of empirically obtained psychometric curves do not perfectly match the ideal observer (Fig. 2) since they asymptote at values that are less than 1 or greater than 0. This is a

well known phenomenon in psychophysics (Wichmann and Hill, 2001), requiring two additional lapse parameters to precisely capture the asymptotes. To account for lapses, we fit a four-parameter psychometric function to the subjects' choice data (Fig. 2a - red, Equation 1 in Methods) with the Palamedes toolbox (Prins and Kingdom, 2018). γ and λ are the lower and upper asymptotes of the psychometric function, which parameterize lapses on low and high rates respectively; ϕ is a sigmoidal function, in our case the cumulative normal distribution; x is the event rate, i.e. the average number of flashes or beeps presented during the one second stimulus period; μ parameterizes the midpoint of the psychometric function and σ describes the inverse slope after correcting for lapses.

How can we be sure that the asymptotes seen in the data truly reflect non-zero asymptotes rather than fitting artifacts or insufficient data at the asymptotes? To test whether lapses were truly necessary to explain the behavior, we fit the curves with and without lapses (Fig. 2b) and tested whether the lapse parameters were warranted. The fit without lapses was rejected in 15/17 rats by the Bayes Information Criterion (BIC), and in all rats by the Akaike Information Criterion (AIC). Fitting a fixed lapse rate across conditions was not sufficient to capture the data, nor was fitting a lapse rate that was constrained to be less than 0.1 (Wichmann and Hill, 2001). Both data pooled across subjects and individual subject data warranted fitting separate lapse rates to each condition ("variable lapses" model outperforms "fixed lapses", "restricted lapses" or "no lapses" in 13/17 individuals based on BIC, all individuals based on AIC and in pooled data based on both, Fig. 2g).

Multisensory trials offer an additional, strong test of ideal observer predictions. In addition to perfect performance on the easiest stimuli, the ideal observer model predicts the minimum

possible perceptual uncertainty achievable on multisensory trials through optimal integration (Ernst and Bulthoff, 2004; Equation 9 in Methods). By definition, better-than-optimal performance is impossible. However, studies in humans, rodents and non-human primates performing multisensory decision-making tasks suggest that in practice, performance occasionally exceeds optimal predictions (Raposo, Sheppard, et al., 2012; Nikbakht et al., 2018; Hou et al., 2018), seeming, at first, to violate the ideal observer model. Moreover, in these datasets, performance on the easiest stimuli was not perfect and asymptotes deviated from 0 and 1. As in these previous studies, when we fit performance without lapses, multisensory performance was significantly supra-optimal (p=0.0012, paired t-test), i.e. better than the ideal observer prediction (Fig. 2c, black points are above the unity line). This was also true when lapse probabilities were assumed to be fixed across conditions (p =0.0018, Fig. 2c purple) or when they were assumed to be less than 0.1 (p=0.0003, Fig. 2c yellow). However, when we allowed lapses to vary freely across conditions, performance was indistinguishable from optimal (Fig. 2d, data points are on the unity line). This reaffirms that proper treatment of lapses is crucial for accurate estimation of perceptual parameters and offers a potential explanation for previous reports of supra-optimality.

Using this improved fitting method, we replicated previous observations (Raposo, Sheppard, et al., 2012; Raposo, Kaufman, and Churchland, 2014) showing that animals have improved sensitivity (lower σ) on multisensory vs. unisensory trials (Fig. 2e, red curve is steeper than green/blue curves; Fig. 2f, top). Interestingly, we observed that animals also had a lower lapse probability ($\lambda + \gamma$) on multisensory trials (Fig. 2e, asymptotes for red curve are closer to 0 and 1; n=17 rats, 347537 trials). This was consistently observed across animals (Fig. 2f bottom, the probability of lapses on

multisensory trials was 0.06 on average, compared to 0.17 on visual, p=1.4e-4 and 0.21 on auditory, p=1.5e-5). We also noticed that compared to unisensory trials, multisensory trials were slightly biased towards high rates. This bias may reflect that animals' decisions do not exclusively depend on the rate of events, but are additionally weakly influenced by the total event count, as has been previously reported on a visual variant of the task (Odoemene et al., 2018).

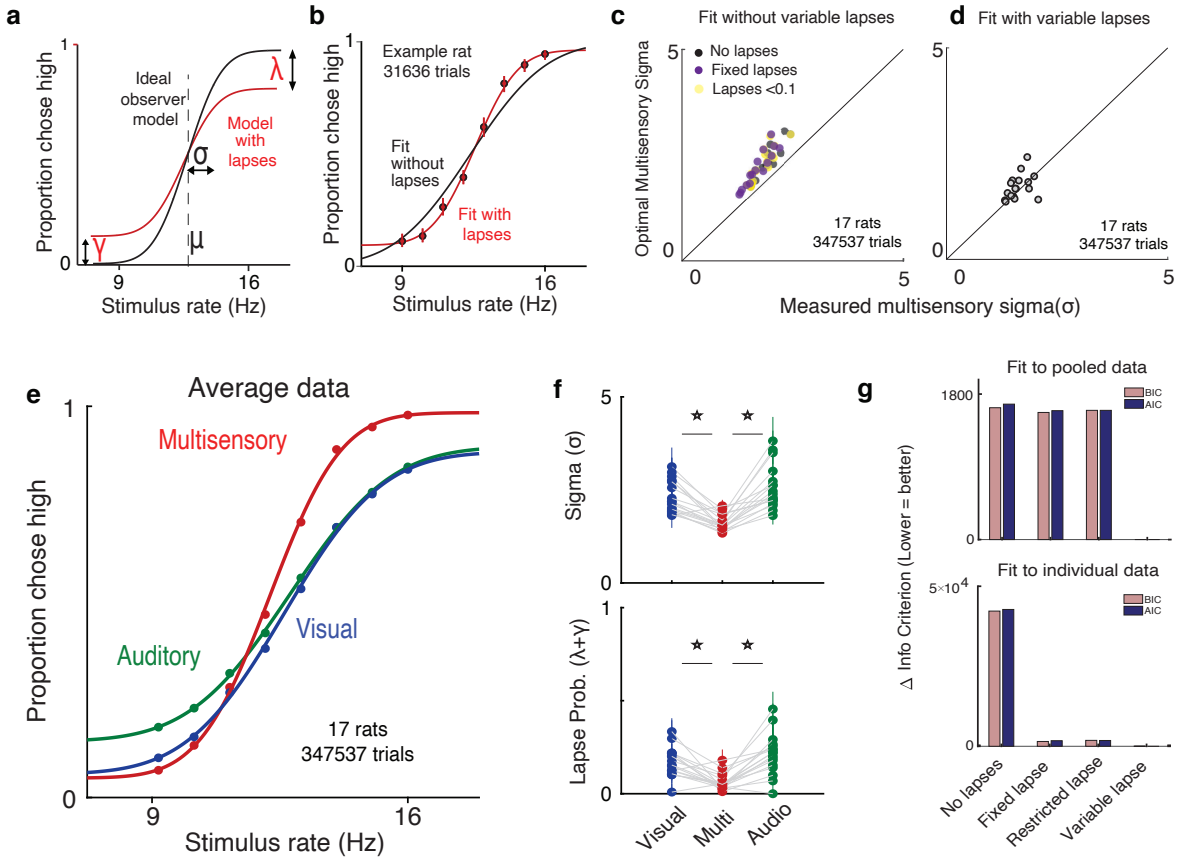


Figure 2 Deviations from ideal observer reflect lapses in judgment. **(a)** Schematic psychometric performance of an ideal observer (black) vs. a model that includes lapses (red). The ideal observer model includes two parameters: midpoint (μ) and inverse slope (σ). The four-parameter model includes μ, σ, and

lapse probabilities for low rate (γ) and high rate choices (λ). Dotted line shows the true category boundary (12.5 Hz). **(b)** Subject data was fit with an two-parameter model without lapses (black) and a four-parameter model with lapses (red). **(c,d)** Ideal observer predictions vs. measured multisensory sigma for fits with and without variable lapses across conditions. **(c)** Multisensory integration seems supra-optimal if lapses are not accounted for (No lapses, black), fixed across conditions (Fixed lapses, purple) or assumed to be less than 0.1 (Restricted lapses, yellow). **(d)** Optimal multisensory integration is restored when allowing lapses to vary freely across conditions. (n = 17 rats. Points represent individual 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. Points represent data pooled across 17 rats, lines represent separate four-parameter fits to each condition. **(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 fits to pooled data across subjects (top) and to individual subject data (bottom). Lower scores indicate better fits. Both metrics favor a model where lapses are allowed to vary freely across conditions ("Variable lapse") over one without lapses ("No lapses"), one with a fixed probability of lapses ("Fixed lapse") or where the lapses are restricted to being less than 0.1 ("Restricted lapse").

Uncertainty-guided exploration offers a novel explanation for lapses where traditional explanations fail

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 accurately captured the

reduction in inverse-slope on multisensory trials, this success does not provide an explanation for why lapses are present in the first place, nor why they differ between stimulus conditions.

To investigate this, we examined possible sources of noise that have traditionally been invoked to explain lapses (Fig. 1d). The first of these explanations is that lapses might be due to a fixed amount of noise added once the decision has been made. These sources of noise could include decision noise due to imprecision (Findling et al., 2018) or motor errors (Wichmann and Hill, 2001). However, these sources should hinder decisions equally across stimulus conditions (Fig. 3-Supplementary Fig. 1b), which cannot explain our observation of condition-dependent lapse rates (Fig. 2f).

A second explanation is that lapses arise due to inattention on a small fraction of trials. Inattention would drive the animal to guess randomly, producing lapse rates whose sum should reflect the probability of not attending (Fig. 3a, Methods). According to this explanation, the lower lapse rate on multisensory trials could reflect increased attention on those trials, perhaps due to their increased bottom-up salience (i.e. two streams of stimuli instead of one). To examine this possibility, we leveraged a multisensory condition that has been used to manipulate perceptual uncertainty without changing salience in rats and humans (Raposo, Sheppard, et al., 2012). Specifically, we interleaved standard matched-rate multisensory trials with "neutral" multisensory trials for which the rate of the auditory stimuli ranged from 9-16 Hz, while the visual stimuli was always 12 Hz. This rate was so close to the category boundary (12.5 Hz) that it did not provide compelling evidence for one choice or the other (Fig. 3d, left), thus reducing the information in the multisensory stimulus and

increasing perceptual uncertainty on "neutral" trials. However, since both "neutral" and "matched" conditions are multisensory, they should be equally salient, and since they are interleaved, the animal would be unable to identify the condition without actually attending to the stimulus. According to the inattention model, matched and neutral trials should have the same rate of lapses, only differing in their inverse-slope σ (Fig. 3a, Fig. 3-Supplementary Fig. 1c).

Contrary to this prediction, we observed higher lapse rates in the "neutral" condition, where trials had higher perceptual uncertainty on average, compared to the "matched" condition (Fig. 3d). This correlation between the average perceptual uncertainty in a condition and its frequency of lapses was reminiscent of the correlation observed while comparing unisensory and multisensory trials (Fig. 2e,f; Fig. 3-Supplementary Fig. 1e).

Having observed that traditional explanations of lapses fail to account for the behavioral observations, we re-examined a key assumption of ideal observer models used in perceptual decision-making - that subjects have complete knowledge about the rules and rewards (Dayan and Daw, 2008). In general, this assumption may not hold true for a number of reasons - even when the stimulus category is known with certainty, subjects might have uncertainty about the values of different actions because they are still in the process of learning (Law and Gold, 2009), because they incorrectly assume that their environment is non-stationary (Yu and Cohen, 2009), or because they forget over time (Gershman, 2015; Drugowitsch and Pouget, 2018). In such situations, rather than always "exploiting" (i.e. picking the action currently assumed to have the highest value), it is advantageous to "explore" (i.e. occasionally pick actions whose value the subject is uncertain

about), in order to gather more information and maximize reward in the long term (Dayan and Daw, 2008). Exploratory choices of the lower value action for the easiest stimuli would resemble lapses, and the sum of lapses would reflect the overall degree of exploration.

Choosing how often to explore is challenging, and requires trading off immediate rewards for potential gains in information - random exploration would reward subjects at chance, but would reduce uncertainty uniformly about the value of all possible stimulus-action pairs, while a greedy policy (i.e. always exploiting) would yield many immediate rewards while leaving lower value stimulus-action pairs highly uncertain (Fig. 3-Supplementary Fig. 2a,b). Policies that explore randomly on a small fraction of trials (e.g. "ϵ-Greedy" policies) do not make prescriptions about how often the subject should explore, and are behaviorally indistinguishable from motor errors when the fraction is fixed (Fig. 3b). One elegant way to automatically balance exploration and exploitation is to explore more often when one is more uncertain about action values. In particular, a form of uncertainty-guided exploration called Thompson sampling is asymptotically optimal in many general environments (Leike et al., 2016), achieving lower regret than other forms of exploration (Fig. 3-Supplementary Fig. 2c). This can be thought of as a dynamic "softmax" policy (Fig. 3c), whose "inverse temperature" parameter (β) scales with uncertainty (Gershman, 2018). This predicts a lower β when values are more uncertain, encouraging more exploration and more frequent lapses, and a higher β when values are more certain, encouraging exploitation. The limiting case of perfect knowledge ($\beta \to \infty$) reduces to the reward-maximizing ideal observer.

Subjects' uncertainty about stimulus-action values is compounded by perceptual uncertainty -

on trials where the stimulus category is not fully known, credit cannot be unambiguously assigned to one stimulus-action pair when rewards are obtained and value uncertainty is only marginally reduced. Hence conditions where trials have higher perceptual uncertainty on average (e.g. unisensory or neutral trials) will have more overlapping value beliefs, encouraging more exploration and giving rise to more frequent lapses (Fig. 3-Supplementary Fig. 2d).

As a result, on neutral multisensory trials, the uncertainty-guided exploration model predicts an increase not only in the inverse slope parameter σ, but also in the rate of lapses, just as we observed (Fig. 3d). In fact, this model predicts that both slope and lapse parameters on neutral trials should match those on auditory trials, since these conditions have comparable levels of perceptual uncertainty. The data was well fit by the exploration model (Fig. 3e, bottom) and satisfied both predictions (Fig. 4-Supplementary source data, Neutral has higher σ and lower β than Multisensory, and matched σ and β to Auditory) . By contrast, the inattention model predicts that both conditions would have the same lapse rates, with the neutral condition simply having a larger inverse slope σ. This model provided a worse fit to the data, particularly missing the data at extreme stimulus values where lapses are most clearly apparent (Fig. 3e, top). Model comparison using BIC and AIC favored the exploration model over the inattention model, both for fits to pooled data across subjects (Fig. 3f top) and fits to individual subject data (Fig. 3f bottom, Fig. 3-Supplementary Fig. 3 , for the 3/5 subjects rejected by ideal observer model i.e. with sizable lapses. Both predictions of the exploration model were confirmed using unconstrained descriptive fits to individuals, and held up for 4/5 subjects)

To further understand the precise relationship between perceptual uncertainty and lapses under this form of exploration, we simulated learning in a Thompson sampling agent for various levels of sensory noise, and found a roughly linear relationship between sensory noise and average lapse rate. Hence we fit a constrained version of the exploration model to the multisensory data from 17 rats, where the degree of exploratory lapses was constrained to be a linear function of that condition's sensory noise (with 2 free parameters - slope and intercept, rather than 3 free parameters for the 3 conditions). This model yielded lower BIC than the unconstrained exploration model in all 14/17 rats that were rejected by the ideal observer model (Fig. 3-Supplementary Fig. 3c), and yielded similar slope and intercept parameters across animals (Fig. 3-Supplementary Fig. 2e).

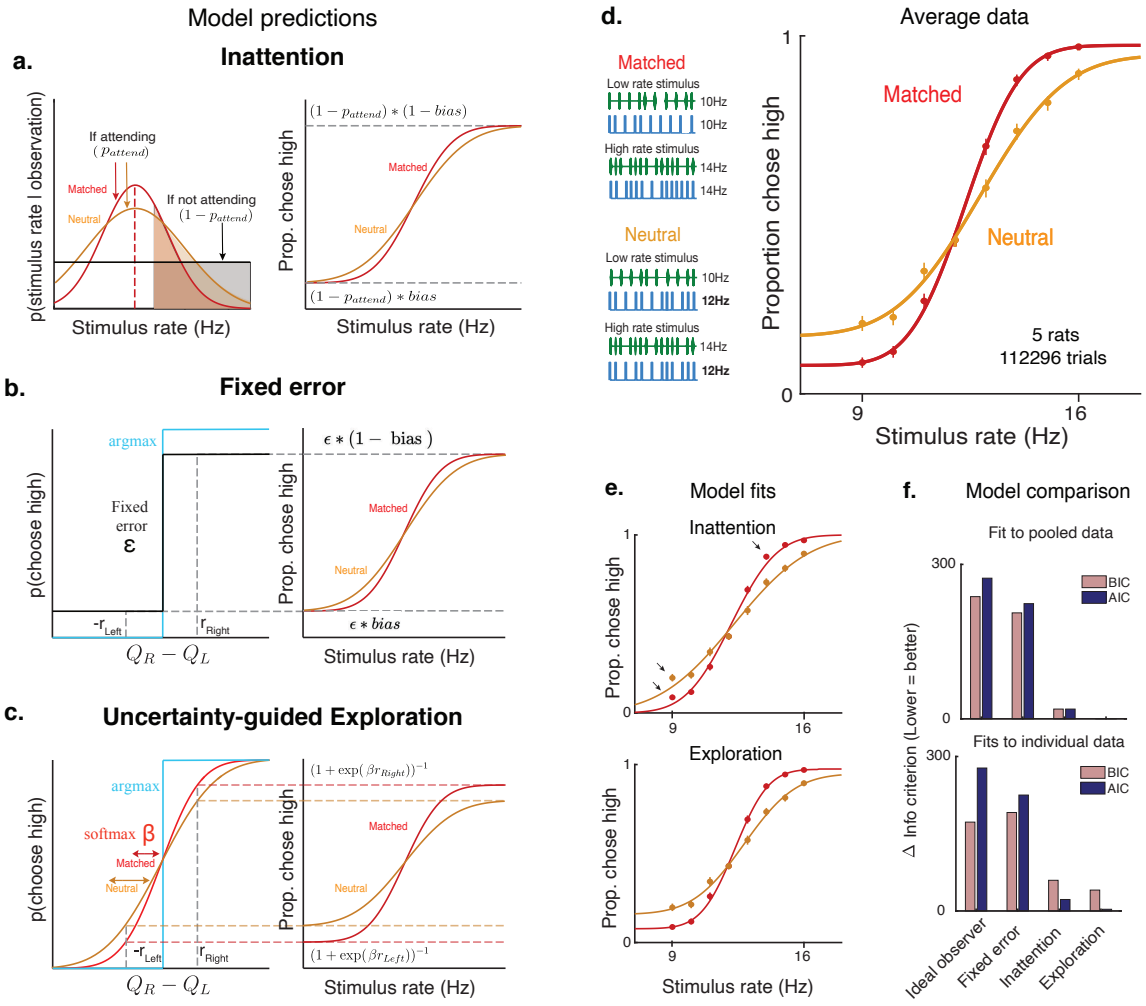


Figure 3 Uncertainty-guided exploration offers a novel explanation for lapses. Models of lapses in decision-making: **(a)** Inattention model of lapses. Left panel: Observer's posterior belief about rate. On a large fraction of trials given by p_{attend}, the observer attends to the stimulus, and has a peaked belief about the rate whose width reflects perceptual uncertainty (red curve on matched trials, orange curve on neutral trials), but on a small fraction of trials given by $1 - p_{attend}$, the observer does not attend to the stimulus (black curve), leading to equal posterior beliefs of rates being high or low (Shaded, clear regions of curves respectively) and guesses according to the probability $bias$, giving rise to lapses (right panel). The sum of

lapse rates then reflects $1 - p_{attend}$, while their ratio reflects the $bias$. Since matched and neutral trials are equally salient, they are expected to have the same p_{attend} and hence similar overall lapse rates. **(b)** Fixed error model of lapses. Lapses could arise due to motor errors ocurring on ϵ fraction of trials, or from decision rules that explore on a fixed proportion ϵ of trials (black), rather than always maximizing reward (blue). The sum of lapses reflects ϵ while their ratio reflects any $bias$ in motor errors or exploration, leading to a fixed rate of lapses across conditions. **(c)** Uncertainty-guided exploration model. Lapses can also arise from more sophisticated exploratory decision rules such as the "softmax" decision rule. Since the difference in expected value from right and left actions ($Q_R - Q_L$) is bounded by the maximum reward magnitudes r_{Right} and r_{Left}, even when the stimulus is very easy, the maximum probability of choosing the higher value option is not 1, giving rise to lapses. Lapse rates on either side are then proportional to the reward magnitude on that side, and to a "temperature" parameter β that is modulated by the uncertainty in action values. Conditions with higher overall perceptual uncertainty (eg. neutral, orange) are expected to have higher value uncertainty, and hence higher lapses. **(d)** Left: multisensory stimuli designed to distinguish between attentional and non-attentional sources of lapses. Standard multisensory stimuli with matched visual and auditory rates (top) and "neutral" stimuli where one modality has a rate very close to the category boundary and is uninformative (bottom). Both stimuli are multisensory and designed to have equal bottom-up salience, and can only be distinguished by attending to them and accumulating evidence. Right: rat performance on interleaved matched (red) and neutral (orange) trials. **(e)** Model fits (solid lines) overlaid on average data points. Deviations from model fits are denoted with arrows. The exploration model (bottom) provides a better fit than the inattention model (top), since it predicts higher lapse rates on neutral trials (orange). **(f)** Model comparison using BIC (pink) and AIC (blue) both favor the uncertainty-guided exploration model for pooled data (top) as well as individual subject

data (bottom).

Reward manipulations confirm predictions of exploration model

One of the key claims of the uncertainty-guided exploration model is that lapses are exploratory choices made with full knowledge of the stimulus, and should therefore depend only on the expected rewards associated with that stimulus category (Fig. 3-Supplementary Fig. 2). This is in stark contrast to the inattention model and many other kinds of disengagement (Fig. 4-Supplementary Fig. 1), according to which lapses are caused by the observer disregarding the stimulus, and hence lapses at the two extreme stimulus levels are influenced by a common underlying guessing process that depends on expected rewards from both stimulus categories. This is also in contrast to fixed error models such as motor error or ϵ-greedy models in which lapses are independent of expected reward (Fig. 3b).

Therefore, a unique prediction of the exploration model is that selectively manipulating expected rewards associated with one of the stimulus categories should only change the explore-exploit tradeoff for that stimulus category, selectively affecting lapses at one extreme of the psychometric function. Conversely, inattention and other kinds of disengagement predict that both lapses should be affected, while fixed error models predict that neither should be affected (Fig. 4a, Fig. 3-Supplementary Fig. 1, Fig. 4-Supplementary Fig. 1).

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.

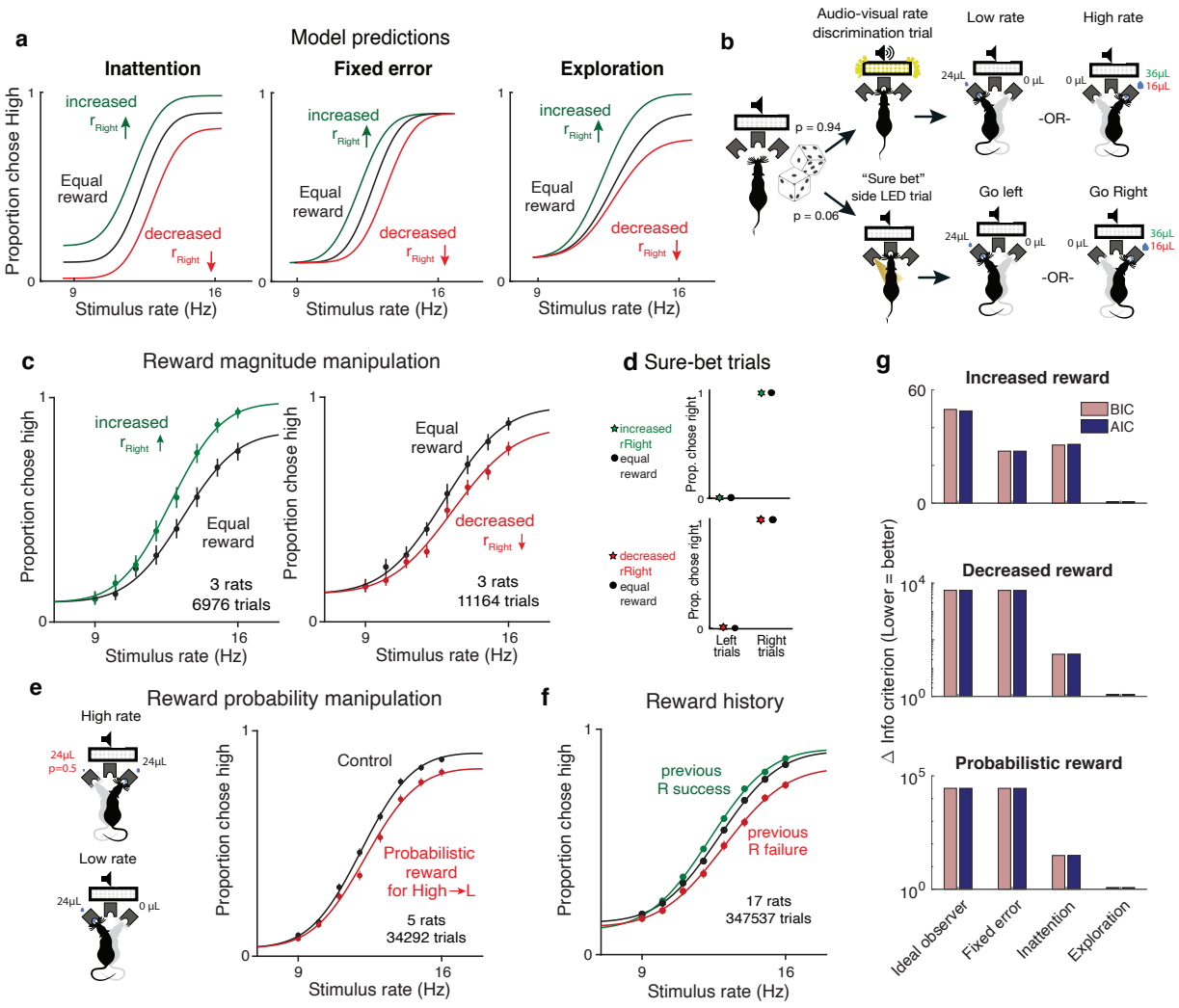


Figure 4 Reward manipulations match predictions of the exploration model. **(a)** The inattention, fixed error and exploration models make different predictions for increases and decreases in the reward magnitude for rightward (high-rate) actions. The inattention model (left panel) predicts changes in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

and low rate choices, while fixed error models such as motor error or ϵ-greedy (center) predict changes in neither lapse, and the uncertainty-dependent exploration model (right) predicts changes in lapses only for high rate choices. Black line denotes equal rewards on both sides; green, increased rightward reward; red, decreased rightward reward. **(b)** Schematic of rate discrimination trials and interleaved "sure bet" trials. The majority of the trials (94%) were rate discrimination trials as described in Figure 1. On sure-bet trials, a pure tone was played during a 0.2 second fixation period and one of the side ports was illuminated once the tone ended to indicate that reward was available there. Rate discrimination and sure-bet trials were randomly interleaved, as were left and right trials, and the rightward reward magnitude was either increased (36 μl) or decreased (16 μl) while maintaining the leftward reward at 24 μl **(c)** Rats' behavior on rate discrimination trials following reward magnitude manipulations. High rate lapses decrease when water reward for high-rate choices is increased (left panel; n=3 rats, 6976 trials), while high-rate lapses increase when reward on that side is decreased (right panel; n=3 rats, 11164 trials). Solid curves are exploration model fits with a single parameter change accounting for the manipulation. **(d)** Rats show nearly perfect performance on sure-bet trials, and are unaffected by reward manipulations on these trials. **(e)** Reward probability manipulation. (Left) Schematic of probabilistic reward trials, incorrect (leftward) choices on high rates were rewarded with a probability of 0.5, and all other rewards were left unchanged. (Right) Rats' behavior and exploration model fits showing a selective increase in high-rate lapses (n=5 rats, 34292 trials). **(f)** Rats' behavior on equal reward trials conditioned on successes (green) or failures (red) on the right on the previous trials resembles effects of reward size manipulations. **(g)** Model comparison showing that AIC and BIC both favor the exploration model on data from all 3 manipulations.

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

ration model. Increasing the reward size on choices associated with high-rates led to a decrease in lapses for the highest rates and no changes in lapses for the lowest rates (Fig. 4c, left; n=3 rats, 6976 trials). Decreasing the reward of choices associated with high-rates led to an increase in lapses for the highest rates and no changes in lapses for the lower rates (Fig. 4c, right; n=3 rats, 11164 trials). This shows that both increasing and decreasing the value of actions associated with one of the stimulus categories selectively affects lapses on that stimulus category, unlike the predictions of the inattention model.

A key claim of the uncertainty-guided exploration model is that the effects of reward manipulations on lapses arise from a selective shift in the trade-off between exploiting the most rewarding action and exploring uncertain ones, rather than from a non-selective bias towards the side with bigger rewards. Importantly, the model predicts that in the absence of uncertainty, decisions should be perfectly exploitative and unaffected by reward imbalances, since subjects would always be comparing perfectly certain, non-zero rewards to zero. To determine whether the effects that we observed were truly driven by uncertainty, we examined performance on randomly interleaved "sure bet" trials on which the uncertainty was very low (Fig. 4b, bottom). On these trials, a pure tone was played during the fixation period, after which an LED at one of the side ports was clearly illuminated, indicating a reward. Sure-bet trials comprised 6% of the total trials, and as with the rate discrimination trials, left and right trials were interleaved. Owing to the low perceptual uncertainty and consequently low value uncertainty, the model predicts that that animals would quickly reach an "exploit" regime, achieving perfect performance on these trials. Importantly, our model predicts that performance on these "sure-bet" trials would be unaffected by imbalances in reward magnitude,

since the "exploit" action remains unchanged.

In keeping with this prediction, performance on sure-bet trials was near perfect (rightward probabilities of 0.003 [0.001,0.01] and 0.989 [0.978,0.995] on go-left and go-right trials respectively), and unaffected following reward manipulations (Fig. 4d: Rightward probabilities of 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 [0.983,0.994] on decreased reward). This suggests that the effects of reward manipulations that we observed (Fig. 4C) are not a default consequence of reward imbalance, but a consequence of a reward-dependent trade-off between exploitation and uncertainty-guided exploration.

As an additional test of the model, we manipulated expected rewards by probabilistically rewarding incorrect choices for one of the stimulus categories. Here, leftward choices on high rate ("go right") trials were rewarded with a probability of 0.5, while leaving all other rewards unchanged (Fig. 4e left). The exploration model predicts that this should selectively increase the value of leftward actions on high rate trials, hence shifting the trade-off towards exploration on high rates and increasing high rate lapses. Indeed, this is what we observed (Fig. 4e right, n=5 animals, 347537 trials), and the effect was strikingly similar to the decreased reward experiment, even though the two manipulations affect high rate action values through changes on opposite actions. This experiment in particular distinguishes the exploration model from motivation-dependent models of disengagement or inattention in which overall reward modulates the total lapse rate through a non-specific process that averages over stimulus categories (Fig. 4-Supplementary Fig. 1 a-c, f). Moreover, this suggests that lapses reflect changes in stimulus-specific action value caused by

changing either reward magnitudes or reward probabilities, as one would expect from the exploration model. Across experiments (Fig. 4-Supplementary source data) and individuals, these changes were captured by selectively changing the relevant baseline action value in the model, despite variability in these baselines.

An added consequence of uncertainty in action values is that it should encourage continued learning even in the absence of explicit reward manipulations. This means that animals should continue to use the outcomes of previous trials to update the values of different actions as long as this uncertainty persists. Such persistent learning has been observed in a number of studies (Busse et al., 2011; Lak et al., 2018; Mendonca et al., 2018; Odoemene et al., 2018; Pinto et al., 2018; Scott et al., 2015). The uncertainty-dependent exploration model predicts that the effect of recent outcome history on action values should manifest as changes in lapse rates, rather than as horizontal biases caused by irrelevant, non-sensory evidence as is often assumed (Busse et al., 2011). For example, the action value of rightward choices should increase following a rightward success, producing similar changes to lapses as increased rightward reward magnitude. As predicted, trials following rewarded and unrewarded rightward choices showed decreased and increased lapses, respectively (Fig. 4f; same rats and trials as in Fig. 2e). Taken together, manipulations of value confirm the predictions of the uncertainty-dependent exploration model (Fig. 4g).

Lapses are a powerful tool for assigning decision-related computations to neural structures based on disruption experiments

The results of the behavioral manipulations (above) predict that unilateral disruption of neural

regions that leads to a one-sided scaling of learnt stimulus-action values should affect lapse rates asymmetrically. In contrast, disruptions to areas that process sensory evidence would lead to horizontal biases without affecting action values or lapses, and disruptions to motor areas that make one of the actions harder to perform irrespective of the stimulus would affect both lapses (Fig. 4-Supplementary Fig. 2a top, middle). Crucially, in the absence of lapses, all three of these disruptions would drive an identical behavioral effect, a horizontal shift of the psychometric function (Fig. 4-Supplementary Fig. 2a bottom). Indeed, the same reward manipulations that gave rise to distinct value biases in rats with sizeable lapses (Fig. 4-Supplementary Fig 2b top) led to horizontal shifts indistinguishable from sensory biases in highly trained rats with negligible lapses on multisensory trials (Fig. 4-Supplementary Fig 2b bottom). This suggests that lapses are actually informative about decision-making computations and can be used as a tool to determine which computations are affected by disruptions of a candidate brain region. To demonstrate this, we identified two candidate areas, secondary motor cortex (M2) and posterior striatum (pStr), that receive convergent input from primary visual and auditory cortices (Fig. 5-Supplementary Fig. 1, results of simultaneous anterograde tracing from V1 and A1; also see Jiang and Kim, 2018; Barthas and Kwan, 2017). In previous work, disruptions of these areas had effects on auditory decisions, including changes in lapses (Erlich et al., 2015; Guo et al., 2018). However, considerable controversy remains as to which computations were affected by those disruptions. The effects were largely interpreted in terms of traditional ideal observer models (see Siniscalchi, H. Wang, and Kwan, 2019 for a notable exception), and thus attributed to perceptual biases (Guo et al., 2018), leaky accumulation (Erlich et al., 2015) or post categorization biases (Piet et al., 2017; Erlich et al.,

2015). Notably, the asymmetric effects on lapses seen in these studies resembled the effects of the reward manipulations in our task, hinting that they may actually arise from action value changes. Importantly, these existing studies used only auditory stimuli, so were limited in their ability to distinguish sensory-specific deficits from action value deficits.

Here, we used analyses of lapses to determine the decision-related computations altered by unilateral disruption of M2 and pStr. If these disruptions affected action values, the exploration model makes three strong predictions. First, because action values are computed late in the decision-making process, the model predicts that the effects should not depend on the modality of the stimulus. We therefore performed disruptions in animals doing interleaved auditory, visual and multisensory trials. If pStr and M2 indeed compute action value, then following unilateral disruption of these areas, our model should capture changes to all three modalities by a single parameter change to the contralateral action value. Second, these disruptions should selectively affect lapses on stimuli associated with contralateral actions, irrespective of the stimulus-response contingency. To test this, we performed disruptions on animals trained on standard and reversed contingencies. Finally, because altered action values should have no effect when there is no uncertainty and consequently no exploration, disruption to pStr and M2 should spare performance on sure-bet trials (Fig. 4b, bottom).

We suppressed activity of neurons in each of these areas using muscimol, a $GABA_A$ agonist, during our multisensory rate discrimination task. We implanted bilateral cannulae in M2 (Fig. 5a, Fig. 5-Supplementary Fig. 2b; n = 5 rats; +2 mm AP 1.3 mm ML, 0.3 mm DV) and pStr (Fig.

5a, Fig. 5-Supplementary Fig. 2a; n = 6 rats; -3.2 mm AP, 5.4 mm ML, 4.1 mm DV). On control days, rats were infused unilaterally with saline, followed by unilateral muscimol infusion the next day (M2: 0.1-0.5 μg, pStr 0.075-0.125 μg). We compared performance on the multisensory rate discrimination task for muscimol days with preceding saline days. Inactivation of the side associated with low-rate choices biased the animals to make more low-rate choices (Fig. 5b; left 6 panels: empty circles, inactivation sessions; full circles, control sessions), while inactivation of the side associated with high-rates biased them to make more high-rate choices (Fig. 5b, right 6 panels). The inactivations largely affected lapses on the stimulus rates associated with contralateral actions, while sparing those associated with ipsilateral actions (Fig. 5c). These results recapitulated previous findings, and were strikingly similar to the effects we observed following reward manipulations (as seen in Fig. 4c, right panel). These effects were seen across areas (Fig. 5b, top, M2; bottom, pStr) and modalities (Fig. 5b; green, auditory; blue, visual and red, multisensory).

Fitting averaged data across rats with the exploration model revealed that, in keeping with the first model prediction, the effects on lapses in all modalities could be captured by scaling the contralateral action value by a single parameter (Fig. 5b, joint fits to control [solid lines] and inactivation trials [dotted lines] across modalities with the "biased value" model, differing only by a single parameter), similar to the reward manipulation experiments. Animals that were inactivated on the side associated with high rates showed increased lapses on low-rate trials (Fig. 5c, bottom right; data points are above the unity line; n=9 rats), but unchanged lapses on high-rate trials (Fig. 5c, top right; data points are on the unity line). This was consistent across areas and modalities (Fig. 5c; M2, triangles; pStr, circles; blue, visual; green, auditory; red, multisensory). Similarly, animals

that were inactivated on the side associated with low rates showed the opposite effect: increased lapses on high-rate trials (Fig. 5c, top left; n=10 rats), while lapses did not change for low-rate trials (Fig. 5c bottom left). Fits to individual animals revealed that the majority of animals were best fit by the "biased value" model (6/8 rats in M2 - Fig. 5-Supplementary Fig. 3, 7/11 in pStr - Fig. 5-Supplementary Fig. 4), and the remaining animals were best fit by the "biased effort" model.

In keeping with the second prediction, when we compared the effects of the disruptions in animals trained on standard and reversed contingencies (low rates rewarded with leftward or rightward actions respectively), the effects were always restricted to lapses on the stimuli associated with the side contralateral to the inactivation (Fig. 5-Supplementary Fig. 5), always resembling a devaluation of contralateral actions (Fig. 5-Supplementary Fig. 6).

A model comparison across rats revealed that a fixed multiplicative scaling of contralateral value captured the inactivation effects much better than a fixed reduction in contralateral sensory evidence or a fixed addition of contralateral motor effort, both for M2 (Fig. 5e top) and pStr (Fig. 5e bottom). In uncertain conditions, this reduced contralateral value gives rise to more exploratory choices and hence more lapses on one side (Fig. 5f top).

The final prediction of the exploration model is that changes in action value will only affect trials in which there was uncertainty about the outcome. In keeping with that prediction, performance was spared on sure-bet trials (Fig. 5d): rats made correct rightward and leftward choices regardless of the side that was inactivated. This observation provides further reassurance that the changes we observed on more uncertain conditions did not simply reflect motor impairments that drove a

tendency to favor ipsilateral movements. Additional movement parameters such as wait time in the center port and movement times to ipsilateral and contralateral reward ports were likewise largely spared (Fig. 5-Supplementary figure 7), suggesting that effects on decision outcome were not due to an inactivation-induced motor impairment.

Together, these results demonstrate that lapses are a powerful tool for interpreting behavioral changes in disruption experiments. For M2 and pStr disruptions, our analysis of lapses and deployment of the exploration model allowed us to reconcile previous inactivation studies. Our results suggest that M2 and pStr have a lateralized, modality-independent role in computing the expected value of actions (Fig. 5f bottom).

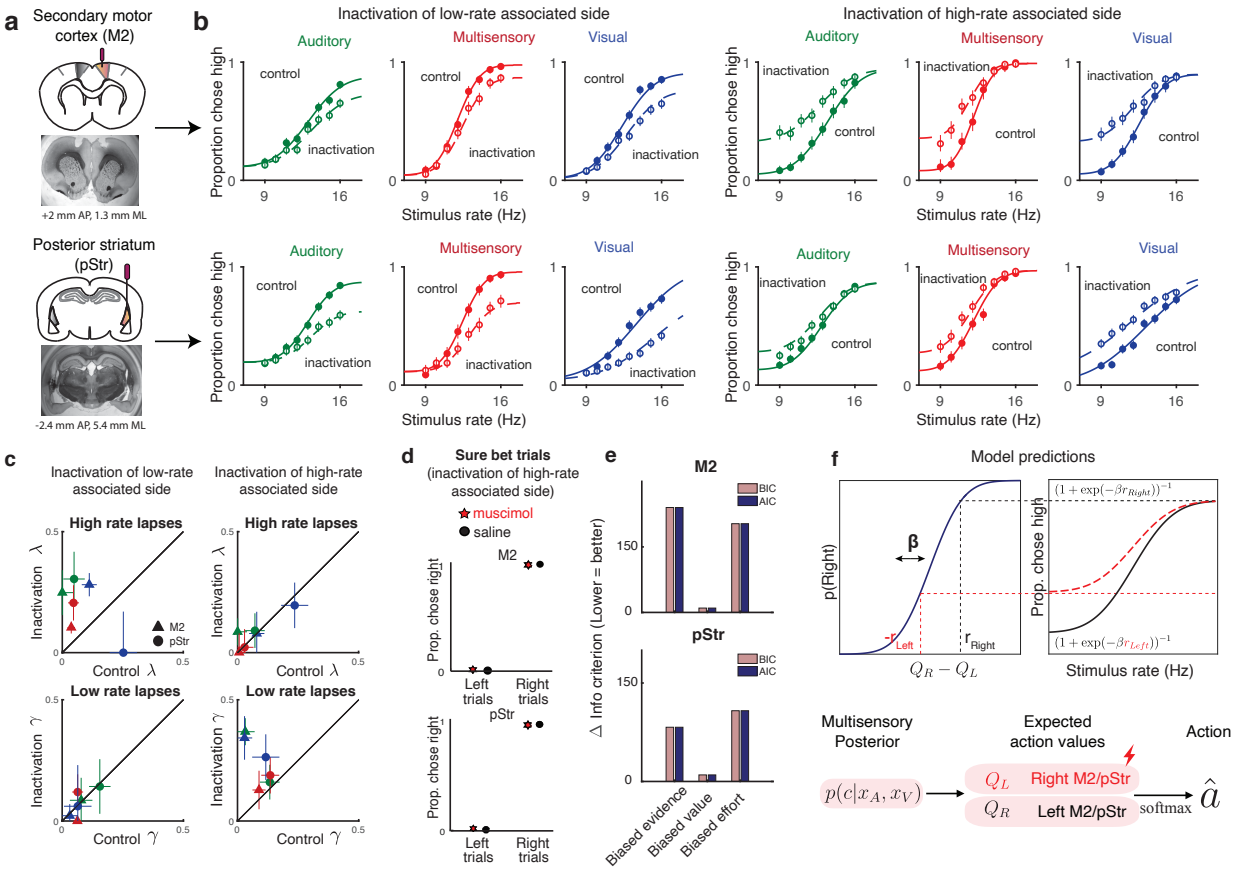


Figure 5 Inactivation of secondary motor cortex and posterior striatum affects lapses, suggesting a role in action value encoding. **(a)** Schematic of cannulae implants in M2 (top) and pStr (bottom) and representative coronal slices. For illustration purposes only, the schematic shows implants in the right hemisphere, however, the inactivations shown in panel (b) were performed unilaterally on both hemispheres. **(b)** Unilateral inactivation of M2 (top) and pStr (bottom). Left 6 plots: inactivation of the side associated with low-rates shows increased lapses for high rates on visual (blue), auditory (green) and multisensory (red) trials (M2: n=5 rats; 10329 control trials, full line; 6174 inactivation trials, dotted line; pStr: n=5 rats; 10419 control trials; 6079 inactivation trials). Right 6 plots: inactivation of the side associated with high-rates shows increased lapses for low rates on visual, auditory and multisensory trials (M2: n=3 rats; 5678 control trials; 3816 inactivation trials; pStr: n=6 rats; 11333 control trials; 6838 inactivation trials). Solid lines are exploration model fits, accounting for inactivation effects across all 3 modalities by scaling all contralateral values by a single parameter. **(c)** Increased high rate lapses following unilateral inactivation of the side associated with low-rates (top left); no change in low rate lapses (bottom left) and vice versa for inactivation of the side associated with high-rates (top, bottom right). Control data on the abscissa is plotted against inactivation data on the ordinate. Same animals as in **b**. Green, auditory trials; blue, visual trials; red, multisensory trials. Abbreviations: posterior striatum (pStr), secondary motor cortex (M2). **(d)** Sure bet trials are unaffected following inactivation. Pooled data shows that rats that were inactivated on the side associated with high rates make near perfect rightward and leftward choices Top, M2 (3 rats); bottom, pStr (6 rats). **(e)** Model comparison of three possible multisensory deficits - reduction of contralateral evidence by a fixed amount (left), reduction of contralateral value by a fixed amount (center), or an increased contralateral effort by a fixed amount (right). Both AIC and BIC suggest a value deficit **(f)** Proposed computational role of M2

and Striatum. Lateralized encoding of left and right action values by right and left M2/pStr (bottom) explains the asymmetric effect of unilateral inactivations on lapses (top).

DISCUSSION

Perceptual decision-makers have long been known to display a small fraction of errors even on easy trials. Until now, these "lapses" were largely regarded as a nuisance and lacked a comprehensive, normative explanation. Here, we propose a novel explanation for lapses: that they reflect a strategic balance between exploiting known rewarding options and exploring uncertain ones. Our model makes strong predictions for lapses under diverse decision-making contexts, which we have tested here. First, the model predicts more lapses on conditions with higher perceptual uncertainty, such as unisensory (Fig. 2) or neutral (Fig. 3), compared to matched multisensory or sure-bet conditions. Second, the model predicts that stimulus-specific reward manipulations should produce stimulus-specific effects on lapses, sparing decisions about un-manipulated or highly certain stimulus-action pairs (Fig. 4). Finally, the model predicts that lapses should be affected by perturbations to brain regions that encode action value. Accordingly, we observed that inactivations of secondary motor cortex and posterior striatum affected lapses similarly across auditory, visual and multisensory decisions, and could be accounted for by a one-parameter change to the action value (Fig. 5). Taken together, our model and experimental data argue strongly that far from being a nuisance, lapses are informative about animals' subjective action values and reflect a trade-off between exploration and exploitation.

Considerations of value have provided many useful insights into aspects of behavior that seem sub-optimal at first glance from the perspective of perceptual ideal observers. For instance, many perceptual tasks are designed with accuracy in mind - defining an ideal observer as one who maximizes accuracy, in line with classical signal detection theory. However, in practice, the success or failure of different actions may be of unequal value to subjects, especially if reward or punishment is delivered explicitly, as is often the case with non-human subjects. This may give rise to biases that can only be explained by an observer that maximizes expected utility (Dayan and Daw, 2008). Similarly, outcomes on a given trial can influence decisions about stimuli on subsequent trials through reinforcement learning, giving rise to serial biases. These biases occur even though the ideal observer should treat the evidence on successive trials as independent (Lak et al., 2018; Mendonca et al., 2018). When subjects can control how long they sample the stimulus, subjects maximizing reward rate may choose to make premature decisions, sacrificing accuracy for speed (Bogacz et al., 2006; Drugowitsch, DeAngelis, et al., 2014). Finally, additional costs of exercising mental effort could lead to bounded optimality through "satisficing" or finding good enough solutions (Mastrogiorgio and Petracca, 2018; Fan, Gold, and Ding, 2018).

Here, we take further inspiration from considerations of value to provide a novel, normative explanation for lapses in perceptual decisions. Our results argue that lapses are not simply accidental errors made as a consequence of attentional "blinks" or motor "slips", but can reflect a deliberate, internal source of behavioral variability that facilitates learning and information gathering when the values of different actions are uncertain. This explanation connects a well known strategy in value-based decision making to a previously mysterious phenomenon in perceptual decision

584 making.

585 Although exploration no longer yields the maximum utility on any given trial, it is critical for
586 environments in which there is uncertainty about expected reward or stimulus-response contingency,
587 especially if these need to be learnt or refined through experience. By encouraging subjects to
588 sample multiple options, exploration can potentially improve subjects' knowledge of the rules of
589 the task, helping them to increase long-term utility. This offers an explanation for the higher rate
590 of lapses seen in humans on tasks with abstract (Raposo, Sheppard, et al., 2012), non-intuitive
591 (Mihali et al., 2018) or non-verbalizable (Flesch et al., 2018) rules.Exploration is also critical for
592 dynamic environments in which rules or rewards drift or change over time. Subjects adapted to such
593 dynamic real-world environments might entertain the possibility of non-stationarity even in tasks
594 or periods where rewards are truly stationary, and such mismatched beliefs predict residual levels
595 of exploration even in well-trained subjects (Fig. 3-Supplementary Fig. 2g middle). Such beliefs
596 could be probed by challenging subjects with unsignalled changes in rewards and measuring how
597 quickly they recover from these change-points. For instance, primates with higher levels of tonic
598 exploration on cognitive set-shifting tasks (Ebitz et al., 2019) are more flexible and make fewer
599 perseverative errors at change-points, at the cost of more lapses in rule adherence during stable
600 periods.

601 Balancing exploration and exploitation is computationally challenging, and the mechanism
602 we propose here, Thompson sampling, is an elegant heuristic for achieving this balance. This
603 strategy has been shown to be utilized by humans in value-based decision making tasks (Wilson

et al., 2014; Speekenbrink and Konstantinidis, 2015; Gershman, 2018) and is asymptotically optimal even in partially observable environments involving perceptual uncertainty such as ours (Fig. 3-Supplementary Fig. 2c, Leike et al., 2016). It can be naturally implemented through a sampling scheme where the subject samples action values from a learnt distribution and then maximizes with respect to the sample. This strategy predicts that conditions with higher perceptual uncertainty and consequently higher value uncertainty should have more exploration, and consequently higher lapse rates, explaining the pattern of lapse rates we observed on unisensory vs. multisensory trials as well as on neutral vs. matched trials. A lower rate of lapses on multisensory trials has also been reported on a visual-tactile task in rats (Nikbakht et al., 2018) and a vestibular integration task in humans (Bertolini et al., 2015) and can potentially account for the apparent supra-optimal integration that has been reported in a number of rodent, non-human primate and human studies (Nikbakht et al., 2018; Hou et al., 2018; Raposo, Sheppard, et al., 2012). A strong prediction of uncertainty guided exploration is that the animal should quickly learn to exploit on conditions with little or no uncertainty, as we observed on sure-bet trials (Fig. 4d, 5d).

Uncertainty-guided exploration also predicts that exploratory choices, and consequently lapses, should decrease with training as the animal becomes more certain of the rules and expected rewards, explaining training-dependent effects on lapses in our rats (Fig. 3-Supplementary Fig 2g right) and similar effects reported in primates (Law and Gold, 2009; Cloherty et al., 2019). This can also potentially explain why children have higher lapse rates (Witton, Talcott, and Henning, 2017; Manning et al., 2018), as they have been shown to be more exploratory in their decisions than adults (Lucas et al., 2014).

A unique prediction of the exploration model is that one-sided reward manipulations should have one-sided effects on lapses, unlike the inattention or motor error models. These predictions are borne out in our data (Fig. 4c), moreover they offer a principled, theoretically grounded way to distinguish between different sources of lapses. This approach can be extended to connect richer statistical descriptions of behavior to psychological variables such as evidence and action value. For instance, some authors have proposed that some of the variance attributed to lapses can be accounted for by allowing psychometric parameters to drift across trials (Roy et al., 2018) or switch between different settings (Ashwood et al., 2019). Whether this parametric non-stationarity arises from non-stationary evidence weighting across trials caused by inattention, variable attention (Shen and Ma, 2019) or attention to irrelevant evidence (Busse et al., 2011), or whether it arises from non-stationary beliefs about action values that encourage continued learning (Lak et al., 2018) and bouts of exploration (Ebitz et al., 2019) can be tested using one-sided reward manipulations, and by extending our model to include trial-by-trial updates of action value based on the history of evidence and outcomes (Pisupati et al., 2019). By decoupling the values of different actions on the two stimulus categories, one-sided reward manipulations distinguish between incorrect decisions made due to a lack of knowledge about the stimulus category (i.e. inattention) and those made despite this knowledge, due to uncertainty about action values (i.e. exploration). An alternative way to decouple these two kinds of errors would be to offer subjects additional actions, for e.g. by adding explicit "opt-out" actions (Zatka-Haas et al., 2019), or by adding task-irrelevant actions that subjects need to learn to avoid (Mihali et al., 2018), affording more opportunities to distinguish exploratory and inattentive decisions than tasks with two alternative actions.

In addition to diagnosing or remedying lapses, the exploration model can be used to harness lapses to pinpoint decision-making computations in the brain. Our model suggests that the asymmetric effects on lapses seen during unilateral inactivations of prefrontal and striatal regions (Fig. 5b) arise from a selective devaluation of learnt contralateral stimulus-action values. This interpretation reconciles a number of studies that have found asymmetric effects of inactivating these areas during perceptual decisions (Erlich et al., 2015; Zatka-Haas et al., 2019; L. Wang et al., 2018; Guo et al., 2018) with their established roles in encoding action value (Sul et al., 2011) during value-based decisions , and strengthens previous proposals that these areas arbitrate between perceptual and value-based influences on decisions.(Lee et al., 2015; Barthas and Kwan, 2017; Siniscalchi, H. Wang, and Kwan, 2019) The effects of inactivation in these studies is consistent with a "devaluation" deficit, or multiplicative scaling of learnt stimulus-action values, resembling the majority of our inactivations (6/8 rats in M2, 7/11 in pStr) and selectively affecting lapses on stimuli strongly associated with the devalued actions. However, inactivations sometimes resembled additive deficits in action value (2/8 rats in M2, 4/11 in pStr), akin to an added "effort" in performing the associated action irrespective of its learnt value, consistent with some reports in striatum (Tai et al., 2012).Further work will be needed to precisely understand the nature of value representations in these regions and why they are sometimes multiplicatively and sometimes additively impacted by inactivations.

An open question that remains is how the brain might tune the degree of exploration in proportion to uncertainty. An intriguing candidate for this is dopamine, whose phasic responses have been shown to reflect state uncertainty (Starkweather et al., 2017; Babayan, Uchida, and Gershman,

2018; Lak et al., 2018), and whose tonic levels have been shown to modulate exploration in mice on a lever-press task (Beeler et al., 2010), and context-dependent song variability in songbirds (Leblois, Wendel, and Perkel, 2010). Dopaminergic genes have been shown to predict individual differences in uncertainty-guided exploration in humans (Frank et al., 2009), and dopaminergic disorders such as Parkinson's disease have been shown to disrupt the uncertainty-dependence of lapses across conditions on a multisensory task (Bertolini et al., 2015), while L-Dopa, a Parkinson's drug and dopamine precursor, has been shown to attentuate uncertainty-guided exploration (Chakroun et al., 2019). 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). Finally, tonic exploration and lapses of rule adherence are reduced in non-human primates that are administered cocaine (Ebitz et al., 2019), which interferes with dopamine transport. 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

Table 1: Key Resources

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
strain, strain background (Rattus norvegicus domestica, male and female)	Long-Evans Rat	Taconic Farms	SimTac:LE	TAC: LONGEV-M, TAC: LONGEV-F
recombinant DNA reagent	AAV2.CB7.CI.EGFP.WPRE.RBG	UPenn Vector Core		Obtained from the laboratory of Dr. Partha Mitra at CSHL
recombinant DNA reagent	AAV2.CAG.tdTomato.WPRE.SV40	UPenn Vector Core		Obtained from the laboratory of Dr. Partha Mitra at CSHL
chemical compound, drug	Muscimol	abcam	ab120094	
software, algorithm	PALAMEDES toolbox	Prins & Kingdom 2018		doi: 10.3389/fpsyg.2018.01250
software, algorithm	MATLAB	The Mathworks, Inc.		

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 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 the visual modality provided evidence that was so close to the category boundary (12 Hz) that it did not support one choice or the other (Raposo, Sheppard, et al., 2012).

To probe the effect of value on lapses, we manipulated either reward magnitude or reward probability associated with high rates, while keeping low rate trials unchanged. To increase or decrease reward magnitude associated with high rates, the amount of water dispensed on the right port was increased or decreased to 36 μl or 16 μl respectively, while the reward on the left port was maintained at 24 μl. To manipulate reward probability, we occasionally rewarded rats on the

(incorrect) left port on high rate trials with a probability of 0.5. The right port was still rewarded with a probability of 1 on high rates, and reward probabilities on low rate trials were unchanged (1 on the left port, 0 on the right).

Analysis of behavioral data.

Psychometric curves. Descriptive four-parameter psychometric functions were fit to choice data using the Palamedes toolbox (Prins and Kingdom, 2018). Psychometric functions were parameterized as:

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

where γ and λ are the lower and upper asymptote of the psychometric function, which parametrize the lapse rates on low and high rates, respectively. ϕ is a cumulative normal function; x is the event rate, i.e. the number of flashes or beeps presented during the one second stimulus period; μ parametrizes the x-value at the midpoint of the psychometric function and σ describes the inverse slope. 95% Confidence intervals on these parameters were generated via bootstrapping based on 1000 simulations.

Our definition of lapses is restricted to strictly *asymptotic* errors following Wichmann and Hill, 2001, and not simply errors on the easiest stimuli tested. Errors on the easiest stimuli could in general arise not just from lapses (strictly defined) but also from perceptual errors caused by low sensitivity to the stimulus, an insufficient stimulus range or non-stationary weights (Busse et al., 2011; Roy et al., 2018). However we do not consider easy errors alone to be evidence of lapses and only consider asymptotic errors. To confirm the necessity of including the lapse parameters, we fit

the following variants of the model above, including lapse parameters when warranted by model comparison using AIC/BIC:

No lapses: This model forces $\lambda = \gamma = 0$ for all conditions (visual, auditory, multisensory) and only allows σ and μ parameters to vary across conditions.

Fixed lapses: This model allows for a fixed λ and γ (which may be unequal) across all conditions.

Restricted lapses: This model allows λ and γ to vary across conditions, but restricts $\lambda + \gamma$ to be less than 0.1. This corresponds to an often used prior over total lapse rates, embodying the belief that lapse trials are infrequent. (Wichmann and Hill, 2001; Prins and Kingdom, 2018)

Variable lapses: This model allows both λ and γ to vary freely across conditions, allowing them each to take any value between 0 and 1 (as long as their sum also lies between 0 and 1).

Modeling

Ideal observer model

We can specify an ideal observer model for our task using Bayesian Decision Theory (Dayan and Daw, 2008). This observer maintains probability distributions over previously experienced stimuli and choices, computes the posterior probability of each action being correct given its observations and picks the action that yields the highest expected reward.

Let the true category on any given trial be c_{true}, the true stimulus rate be s_{true} and the animal's noisy visual and auditory observations of s_{true} be x_V and x_A, respectively. We assume that the two sensory channels are corrupted by independent gaussian noise with standard deviation σ_A and σ_V,

761 respectively, giving rise to conditionally independent observations.

$$
\begin{aligned}
p(x_A|s_{true}) = \mathcal{N}(s_{true}, \sigma_A), \quad p(x_V|s_{true}) = \mathcal{N}(s_{true}, \sigma_V), \\
p(x_A, x_V|s_{true}) = p(x_A|s_{true})p(x_V|s_{true})
\end{aligned}
\tag{2}
$$

762 The ideal observer can use this knowledge to compute the likelihood of seeing the current trial's
763 observations as a function of the hypothesized stimulus rate s. This likelihood $\mathcal{L}$ is a gaussian
764 function of s with a mean given by a weighted sum of the observations x_A and x_V,:

$$
\begin{aligned}
\mathcal{L}(s) &= p(x_A, x_V|s) = p(x_A|s)p(x_V|s) \\
&\propto \mathcal{N}(\mu_M, \sigma_M) \\
\mu_M &= w_A x_A + w_V x_V \\
\sigma_M &= (\sigma_A^{-2} + \sigma_V^{-2})^{-\frac{1}{2}} \\
w_A &= \frac{\sigma_M^2}{\sigma_A^2}, \quad w_V = \frac{\sigma_M^2}{\sigma_V^2}
\end{aligned}
\tag{3}
$$

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

$s < \mu_0$. Then,

$$
\begin{aligned}
\mathcal{L}(c = \text{High}) &= p(x_A, x_V | c = \text{High}) \\
&= \int_s p(x_A, x_V, s | c = \text{High}) ds \\
&= \int_s p(x_A, x_V | s) p(s | c = \text{High}) ds \quad \because x_A, x_V \perp c | s \\
&= \int_{s > \mu_0} p(x_A, x_V | s) ds \\
&\propto 1 - \Phi(\mu_0; \mu_M, \sigma_M)
\end{aligned}
\tag{4}
$$

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 probability.

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

where p_{High} and p_{Low} are the prior probabilities of high and low rates respectively. The expected value $Q(a)$ of choosing right or left actions (also known as the action values) is obtained by marginalizing the learnt value of state-action pairs $q(c, a)$ over the unobserved state c.

$$
\begin{aligned}
Q(a = R) &= p(\text{High} | x_A, x_V) q(High, R) + p(\text{Low} | x_A, x_V) q(Low, R) \\
Q(a = L) &= p(\text{High} | x_A, x_V) q(High, L) + p(\text{Low} | x_A, x_V) q(Low, L)
\end{aligned}
\tag{6}
$$

Under the standard contingency, high rates are rewarded on the right and low rates on the left, so for a trained observer that has fully learnt the contingency, $q(High, R) \to r_R, q(High, L) \to 0, q(Low, R) \to 0, q(Low, L) \to r_L$, with r_R and r_L being reward magnitudes for rightward and leftward actions. This simplifies the action values to:

$$\begin{aligned} Q(R) &= p(\text{High}|x_A, x_V) r_R \propto p_{High}(1 - \Phi(\mu_0; \mu_M, \sigma_M)) r_R \\ Q(L) &= p(\text{Low}|x_A, x_V) r_L \propto p_{Low} \Phi(\mu_0; \mu_M, \sigma_M) r_L \end{aligned} \tag{7}$$

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

$$\begin{aligned} \hat{a} &= \text{argmax} Q(a) \\ \text{i.e. } \hat{a} = R &\iff Q(R) > Q(L) \\ &\iff p_{High}(1 - \Phi(\mu_0; \mu_M, \sigma_M)) r_R > p_{Low} \Phi(\mu_0; \mu_M, \sigma_M)) r_L \\ &\iff \Phi(\mu_M; \mu_0, \sigma_M)) > \frac{1}{1 + \frac{p_{High} r_R}{p_{Low} r_L}} \\ &\iff w_A x_A + w_V x_V > \Phi^{-1}(\frac{1}{1 + \frac{p_{High} r_R}{p_{Low} r_L}}; \mu_0, (\sigma_A^{-2} + \sigma_V^{-2})^{-\frac{1}{2}}) \end{aligned} \tag{8}$$

In the special case of equal rewards and uniform stimulus and category priors, this reduces to choosing right when the weighted sum of observations is to the right of the true category boundary, i.e. $w_A x_A + w_V x_V > \mu_0$. Note that this is a deterministic decision rule for any given observations x_A and x_V, however, since these are noisy and gaussian distributed around the true stimulus rate s_{true}, the likelihood of making a rightward decision is given by the cumulative gaussian function Φ:

$$
\begin{aligned}
&\text{For} \quad p_{High} = p_{Low}, r_R = r_L \\
&p(\hat{a} = R|s) = p(w_A x_A + w_V x_V > \mu_0|s) \\
&\qquad = \Phi(s_{true}; \mu_0, \sigma) \\
&\sigma = \begin{cases} \sigma_A \text{ on auditory trials} \\ \sigma_V \text{ on visual trials} \\ (\sigma_A^{-2} + \sigma_V^{-2})^{\frac{1}{2}} \text{ on multisensory trials} \end{cases}
\end{aligned} \tag{9}
$$

We can measure this probability empirically through the psychometric curve. Fitting it with a two parameter cumulative gaussian function yields μ and σ which can be compared to ideal observer predictions. The σ parameter is then taken to reflect sensory noise; and with the assumption of uniform priors and equal rewards, the μ parameter is taken to reflect the subjective category boundary. For the purpose of assessing optimality of integration, σ was individually fit to each condition and compared to ideal observer predictions, but for the purpose of comparing theoretical models of lapses, σ on multisensory conditions was constrained to be optimal for all models. Although μ should equal μ_0 for the ideal observer, in practice it is treated as a free parameter in all models, and deviations of μ from μ_0 could reflect any of three possible suboptimalities: 1) a subjective category boundary mismatched to the true one, possibly arising from the use of irrelevant features such as total event count (Odoemene et al., 2018), 2) mismatched priors, or 3) unequal subjective rewards r_R and r_L of the two actions.

Inattention model

The traditional model for lapse rates assumes that on a fixed proportion of trials, the animal fails to pay attention to the stimulus, guessing randomly between the two actions. We can incorporate this suboptimality into the ideal observer above as follows: Let the probability of attending be p_{attend}. Then, on $1 - p_{attend}$ fraction of trials, the animal does not attend to the stimulus (i.e. receives no evidence), effectively making $\sigma_{sensory} \to \infty$ and giving rise to a posterior that is equal to the prior. On these trials, the animal may choose to maximize this prior (always picking the option that's more likely a-priori, guessing with 50-50 probability if both options are equally likely), or probability-match the prior (guessing in proportion to its prior). Let us call this guessing probability p_{bias}. Then, the probability of a rightward decision is given by marginalizing over the attentional state:

$$\begin{aligned} p(\hat{a} = R|s) &= p(\hat{a} = R|s, \text{attend})p(\text{attend}) + p(\hat{a} = R|s, \sim \text{attend})p(\sim \text{attend}) \\ &= p(\hat{a} = R|s)p_{attend} + p_{bias}(1 - p_{attend}) \end{aligned} \tag{10}$$

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

$$\begin{aligned} p(\hat{a} = R|s_{true}) &= \gamma + (1 - \gamma - \lambda)\Phi(s_{true}; \mu_0, \sigma) \\ \implies \gamma + \lambda &= 1 - p_{attend}, \quad \frac{\gamma}{\gamma + \lambda} = p_{bias} \end{aligned} \tag{11}$$

where γ and λ are the lower and upper asymptotes respectively, collectively known as "lapses". In this model, the sum of the two lapses depends on the probability of attending, which could be modulated in a bottom up fashion by the salience of the stimulus; their ratio depends on the guessing

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

Motor error/ϵ greedy model

Lapses can also occur if the observer doesn't always pick the reward-maximizing or "exploit" decision. This might occur due to random errors in motor execution on a small fraction of trials given by ϵ, or it might reflect a deliberate propensity to occasionally make random "exploratory" choices to gather information about rules and rewards. This is known as an ϵ-greedy decision rule, where the observer chooses randomly (or according to p_{bias}) on ϵ fraction of trials. Both these models yield predictions similar to those of the inattention model:

$$\begin{aligned} & p(\hat{a} = R|s) = p(\hat{a} = R|s)(1-\epsilon) + \epsilon p_{bias} \\ \implies & \gamma + \lambda = \epsilon, \quad \frac{\gamma}{\gamma + \lambda} = p_{bias} \end{aligned} \tag{12}$$

Uncertainty guided exploration model

A more sophisticated form of exploration is the "softmax" decision rule, which explores options in proportion to their expected rewards, allowing for a balance between exploration and exploitation through the tuning of a parameter β known as inverse temperature. In particular, in conditions of greater uncertainty about rules or rewards, it is advantageous to be more exploratory and have a lower β. This form of uncertainty-guided exploration is known as Thompson sampling. It can be implemented by sampling from a belief distribution over expected rewards and maximizing with respect to the sample, reducing to a softmax rule whose β depends on the total uncertainty in

expected reward (Gershman, 2018).

$$
\begin{aligned}
p(\hat{a} = R|Q(a)) &= \frac{\exp \beta Q(R)}{\exp \beta Q(L) + \exp \beta Q(R)} \\
&= \frac{1}{1 + \exp(-\beta(Q(R) - Q(L)))}
\end{aligned}
\tag{13}
$$

The proportion of rightward choices conditioned on the true stimulus rate is then obtained by marginalizing over the latent action values $Q(a)$, using the fact that the choice depends on s only through its effect on $Q(a)$, where ρ is the animal's posterior belief in a high rate stimulus, i.e. $\rho = p(c = High|x_A, x_V)$. ρ is often referred to as the *belief state* in reinforcement learning problems involving partial observability such as our task.

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

Since lapses are the asymptotic probabilities of the lesser rewarding action at extremely easy stimulus rates, we can derive them from this expression by setting $\rho \to 1$ or $\rho \to 0$. This yields

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

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

rightward or leftward rewards, respectively. In practice since β can only be specified up to an arbitrary scaling of reward magnitudes, we either fix r_L=1 and fit β and a reward bias $\frac{r_R}{r_L}$ in units of r_L (for conditions with different expected β), or fix $\beta = 1$ and fit r_L and r_R in units of β (for conditions with the same β where one of the rewards is expected to change).

Such a softmax decision rule has been used to account for suboptimalities in value based decisions (Dayan and Daw, 2008), however it has not been used to account for lapses in perceptual decisions. Other suboptimal decision rules described in perceptual decisions, such as generalized probability matching or posterior sampling (Acerbi, Vijayakumar, and Wolpert, 2014; Drugowitsch, Wyart, et al., 2016; Ortega and Braun, 2013) amount to a softmax on log-posteriors or log-expected values, rather than on expected values, and do not produce lapses since in these decision rules, when the posterior probability goes to 1, so does the decision probability: .

$$p(\hat{a} = R|Q(a)) = \frac{\exp \beta \log Q_R}{\exp \beta \log Q_L + \exp \beta \log Q_R} = \frac{Q_R^\beta}{Q_L^\beta + Q_R^\beta} \Rightarrow \begin{cases} \rho \to 1 \Rightarrow p(R) \to 1 \\ \\ \rho \to 0 \Rightarrow p(R) \to 0 \end{cases} \tag{16}$$

Inactivation modeling

Inactivations were modeled using the following 1-parameter perturbations to the decision making process, while keeping all other parameters fixed:

Biased evidence: A fixed amount of evidence was added to all modalities. This corresponds to adding a rate bias of $K * \sigma_i$ for a condition with sensory noise σ_i with $K > 0$ fixed across modalities, leading to bigger biases for conditions with higher sensory noise.

Biased value: The expected values of one of the actions was scaled down by a fixed factor of $K < 1$ across all modalities. For instance, $Q_{Li} \rightarrow K * Q_{Li}$ produced a rightward biased value for a condition with baseline leftward expected value Q_{Li}. This led to a stimulus-dependent bias in action value and consequently lapses, since Q_{Li} is large and heavily affected for low rate trials, and close to zero and largely unaffected for high rate trials.

Biased effort: A fixed "effort" cost (i.e. negative value) $K < 0$ was added to the expected values of one of the actions for all modalities. This added a stimulus-independent bias in action values, since the difference in expected values was biased away from the effortful action by the same amount irrespective of the stimulus rate.

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. For comparisons of fits to data pooled across subjects, AIC/BIC values were computed with respect to the best fit model, so that the best model had an AIC/BIC of 0. For comparisons of fits to individual subject data, AIC/BIC values for each subject were computed with respect to the best fit model for each subject, so that the best model for that subject had an AIC/BIC of 0, and then summed across subjects.

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 ($\sim$5mm x $\sim$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, 300μm 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 μl/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

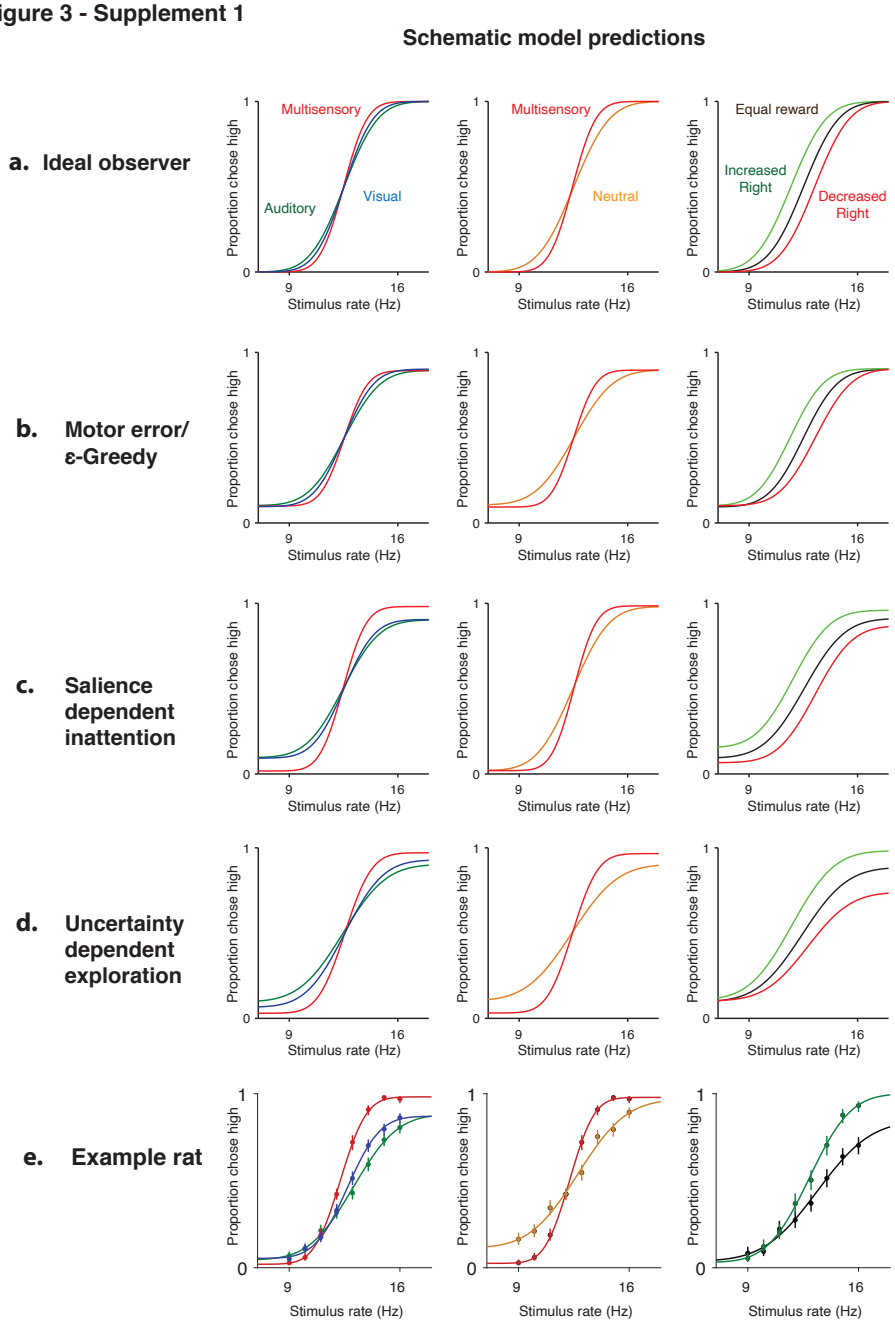
was replaced with sterile saline.

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 μm coronal sections were cut on a vibratome (Leica) and imaged.

929

Figure 3 - Supplement 1



930

931 **Figure 3 Supplement 1: Uncertainty-dependent exploration is the only model that accounts for be-**

932 **havioral data from all three manipulations** Columns: data/predictions for three experimental manipula-

933 tions. Left: unisensory (blue, green) vs. multisensory (red). Middle: matched (red) vs. neutral (orange)
934 multisensory. Right: Increased (green) or decreased (red) rightward reward vs. equal reward (black)
935 on auditory trials. a-d: Four candidate models. (a) Ideal observer model predicts no lapses and only
936 changes in sensitivity/bias across conditions. (b) Fixed motor error model predicts a constant rate of
937 lapses across conditions in addition to changes in sensitivity/bias predicted from the ideal observer. (c)
938 Inattention model predicts that the overall lapse rate (sum of lapses on both sides) depends on the level
939 of bottom-up attentional salience, allowing for different rates for unisensory and multisensory trials. It
940 also predicts that the lapse rate on neutral trials should be equal to that on multisensory trials, and that
941 manipulating rightward reward should affect both lapse rates. (d) Uncertainty-dependent exploration model
942 predicts that overall lapse rate depends on the level of exploratoriness and hence uncertainty associated
943 with that condition, allowing for different lapse rates on unisensory and multisensory trials. It also predicts
944 that the lapse rate on neutral trials should be equal to that on auditory trials and manipulating rightward
945 reward should only affect high rate lapses. (e) Data from an example rat on all three manipulations.

Figure 3 - Supplement 2

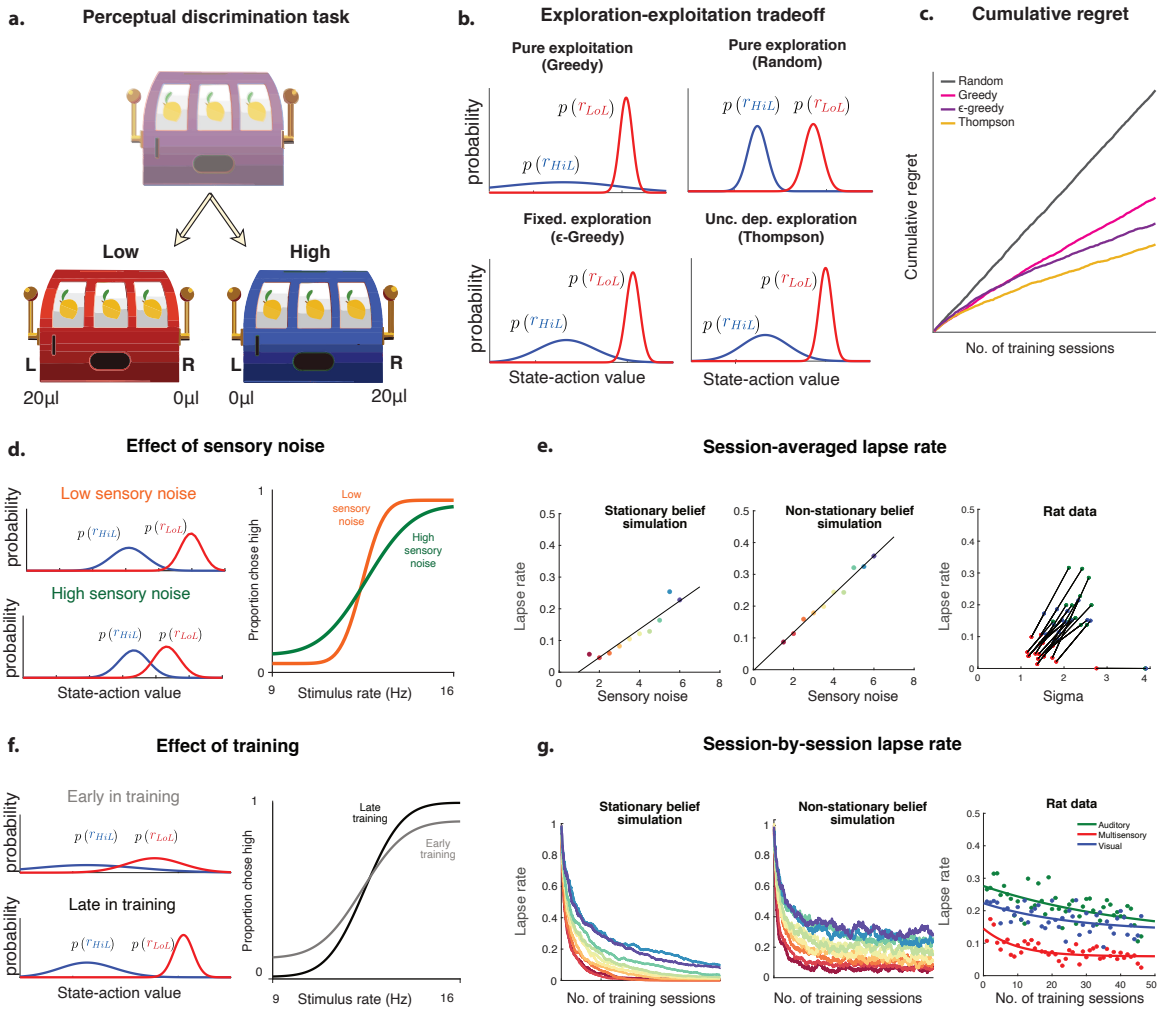


Figure 3 Supplement 2: Thompson sampling, which balances exploration and exploitation, predicts lapses that increase with perceptual noise Schematic illustrating the explore-exploit tradeoff in perceptual two-alternative tasks. (a) Formulation of perceptual decision making task as a partially observable contextual bandit. To solve this task, an observer needs to infer the true category of the stimulus (Low or High) based on noisy observations, and pick the best action given the inferred category (Left for Low, Right for High).

This requires accurately learning the expected rewards from all 4 state-action pairs. (b) Leftward state-action value beliefs i.e. expected reward from leftward actions (L) performed in different states (Hi, Lo) showing different levels of uncertainty depending on policy. Beliefs are updated based on outcomes using a Bayesian update rule that takes into account uncertainty in state estimation. A greedy policy (top left) that always picks the best action maximizes reward and learns well about the preferred state-action pairs (i.e. Lo-L) but has high uncertainty about the non-preferred pairs (Hi-L). A random policy (top right) earns reward at chance, but learns equally well about all state-action pairs. An ϵ-greedy policy (bottom left) learns well about the non-preferred pair, but leaves the choice of ϵ unspecified, and continues exploring even after it has learnt the values well, continuing to forego rewards. Thompson sampling (bottom right) tunes the amount of exploration to the current uncertainties in each value, and balances immediately reward-maximizing decisions with decisions that reduce uncertainty, maximizing average reward in the long term. (c) Cumulative regret i.e. foregone reward accrued by different policies on the rate discrimination task as a function of training, with lower regret being more desirable. Black - random exploration, Pink - greedy, Purple - ϵ-greedy and Yellow-Thompson sampling. Thompson sampling outperforms all other policies, by achieving the minimum regret (d) Learnt beliefs about expected reward with Thompson sampling at various levels of perceptual uncertainty. Low levels of sensory noise (left top) produce more separable beliefs, while higher levels of sensory noise (left bottom) lead to large perceptual uncertainty, yielding highly overlapping belief distributions owing to a reduced ability to assign obtained rewards to one of the states. (right) Simulated performance averaged across 2000 trials of the Bayesian observer, under a Thompson sampling policy. The observer makes fewer exploratory choices for lower levels of sensory noise (orange) owing to the more separable value beliefs, giving rise to lower lapse rates. (e) Session-averaged lapse rates as a function of sensory noise in simulations

(left, center) and multisensory rat data (right). Simulations were done under increasing levels of sensory noise (colors going from hot to cold) under beliefs that action values are stationary (left) or non-stationary (center), solid lines indicate linear best-fit. Individual rat data was fit with a constrained version of the exploration model where total lapse rate was constrained to be linearly related to sensory noise across all modality conditions (auditory - green, multisensory - red, visual - blue). Lines indicate best fit linear constraints for each rat. (f) Learnt beliefs about expected reward with Thompson sampling during early (left top) and late (left bottom) stages of training. Training reduces uncertainty about expected rewards, producing more separable beliefs and yielding less exploration and lower lapse rates over time (right - simulated average performance). (g) Session-wise lapse rates in simulated (left, center) and rat data (right) as a function of both training and sensory noise. Simulations show decreasing lapse rates over training that asymptote at zero under stationary beliefs (left) and to non-zero values dictated by sensory noise under non-stationary beliefs (center). Rat data was separated by session starting from the earliest day of training with all 3 modalities, and combined across rats to produce session-wise fits, and the resulting lapse rates were fit with an exponential curve for each modality (solid lines indicate best-fit curves for multisensory - red, visual-blue, auditory - green))

Figure 3 - Supplement 3

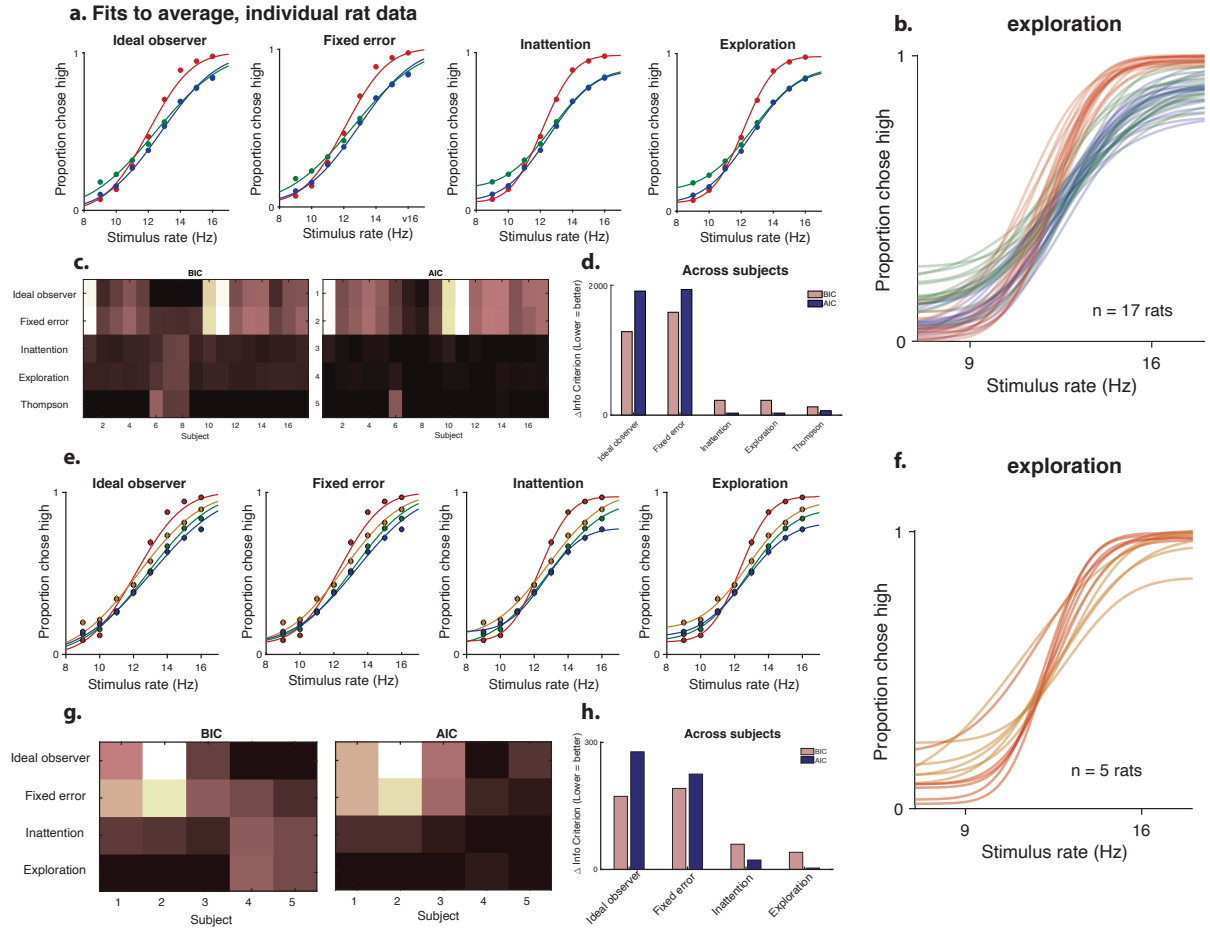


Figure 3 Supplement 3: Uncertainty guided exploration outperforms competing models for average and individual data (a) Fits of the four models (ideal observer, fixed motor error, inattention and exploration) to average rat data on unisensory (blue-visual, green-auditory) and multisensory (red) trials. (b) Exploration model fits to unisensory and multisensory data for 17 individual animals (c) Model comparison for individual animals using BIC (left), AIC (right) of the four aforementioned models, plus a constrained version of the exploration model corresponding to Thompson sampling. Darker colors are lower BICs/AICs, denoting a better fit. (d) Summed model comparison metrics across animals, showing that inattention and exploration models fit the data equally well, and much better than the ideal observer or fixed error models. Thompson

sampling is preferred by BIC, since it fits as well as exploration model but with fewer effective parameters (e) Fits of the four models to average data including neutral trials (orange) provide a stronger test of the inattention model. (f) Exploration model fits to multisensory data including neutral trials for 5 individual animals. (g) Model comparison for individual animals. (h) Summed model comparison metrics across animals shows that the uncertainty-guided exploration model performs better than other models.

Figure 4 - Supplement 1

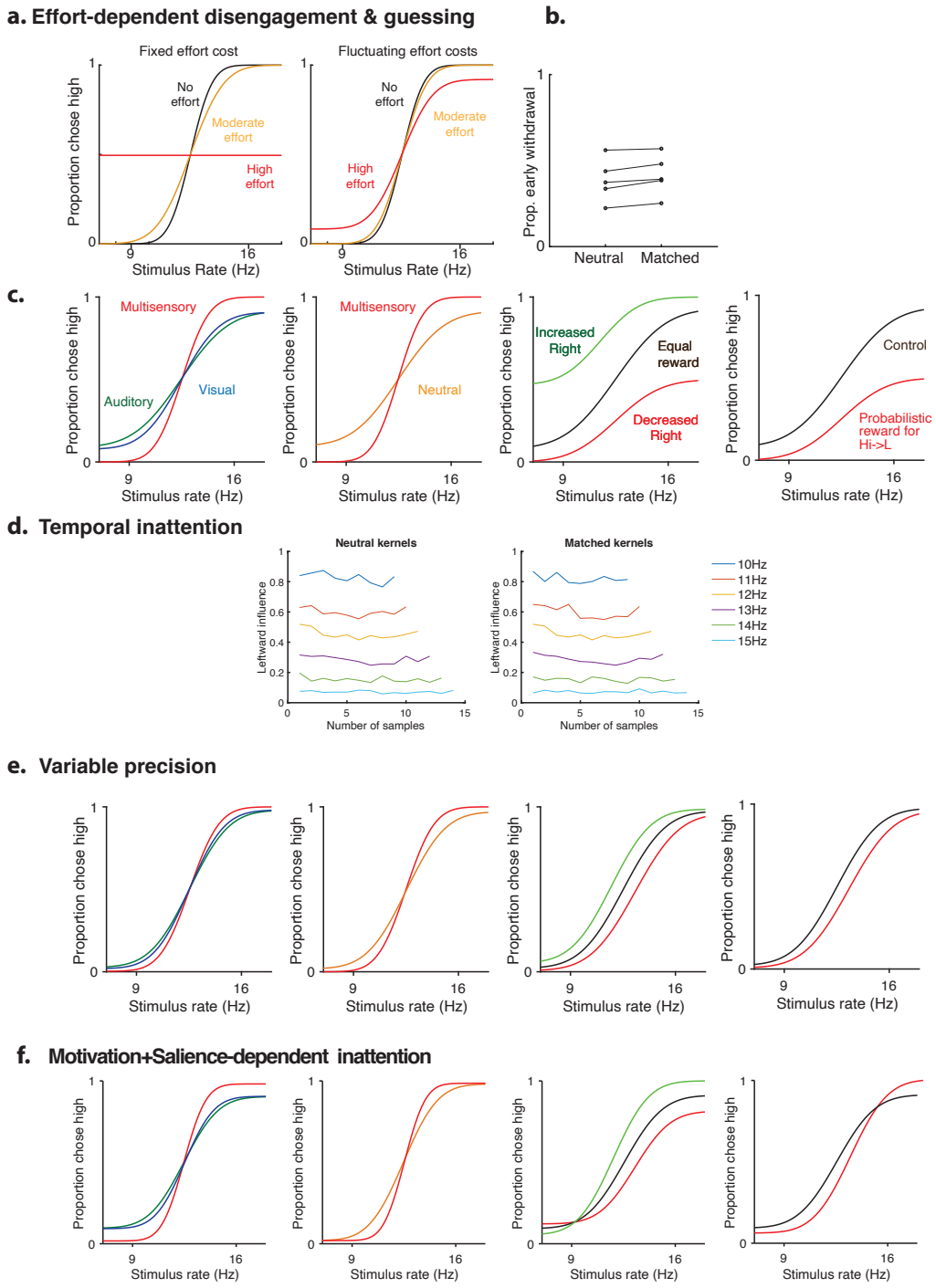


Figure 4 Supplement 1: Alternative models of inattentional lapses. Predictions of alternative models of lapses. (a) Effort-dependent disengagement model: In this model, there is an additional cost or mental effort

to being engaged in the task which could vary with condition, and an additional random guessing action. If the net payoff of engagement is not greater than the average value of a guess, then it guesses randomly. Such a model does not produce lapses if the effort is fixed across trials (left), but could produce lapses if the effort fluctuates from trial to trial (center). (b) Proportion of trials on which the animal withdrew prematurely doesn't vary between matched and neutral trials, suggesting that rats are not disengaging preferentially on neutral trials. (c) Predictions of the effort-dependent disengagement model. The model accurately predicts increased lapses on unisensory trials (left panel, green/blue traces) and neutral multisensory trials (middle left panel, orange trace). However, for asymmetric reward manipulations (middle right - reward magnitude, right - reward probability), the model fails to predict our behavioral observation (Fig. 4d) that only lapses on the manipulated side are affected. (d) Temporal inattention model: in this model, temporal weighting of evidence differs between matched and neutral trials. To test this, we compared psychophysical kernels on matched and neutral trials. The temporal dynamics of attention are unchanged between the two kinds of trials, arguing against the temporal inattention model. (e) Variable precision model: in this model, the sensory noise (or its inverse, precision) fluctuates from trial to trial, producing heavy tailed performance curves with apparent "lapses". The model accurately predicts increased apparent lapses on unisensory trials (left panel, green/blue traces) and neutral multisensory trials (middle left panel, orange trace). However, for asymmetric reward manipulations (middle right, right), the model fails to predict our behavioral observation (Fig. 4d) that lapses only on the manipulated side are affected. Like other models of inattention, it predicts that manipulating reward on one side should affect both lapses. (f) Motivation+salience-dependent inattention: in this model, inattention is determined not just by salience, but also motivation, which in turn depends on average reward. This model's predictions on unisensory, multisensory (left) and neutral (middle left) trials are identical to the

inattention model, but on asymmetric reward manipulations, it predicts that total lapse rate should change as a function of total reward. As a result, when reward magnitude on one side is increased or decreased (middle right), total lapse rate also increases or decreases, in addition to the vertical shifts predicted by inattention. However on the reward probability manipulation (right), it predicts a *decrease* in total lapse rate owing to the overall higher average reward, in addition to a downward shift predicted by inattention, unlike the rat data (Fig. 4e) where overall lapse rate *increases* as a consequence of high rate lapses selectively *increasing*.

Figure 4 - Supplement 2

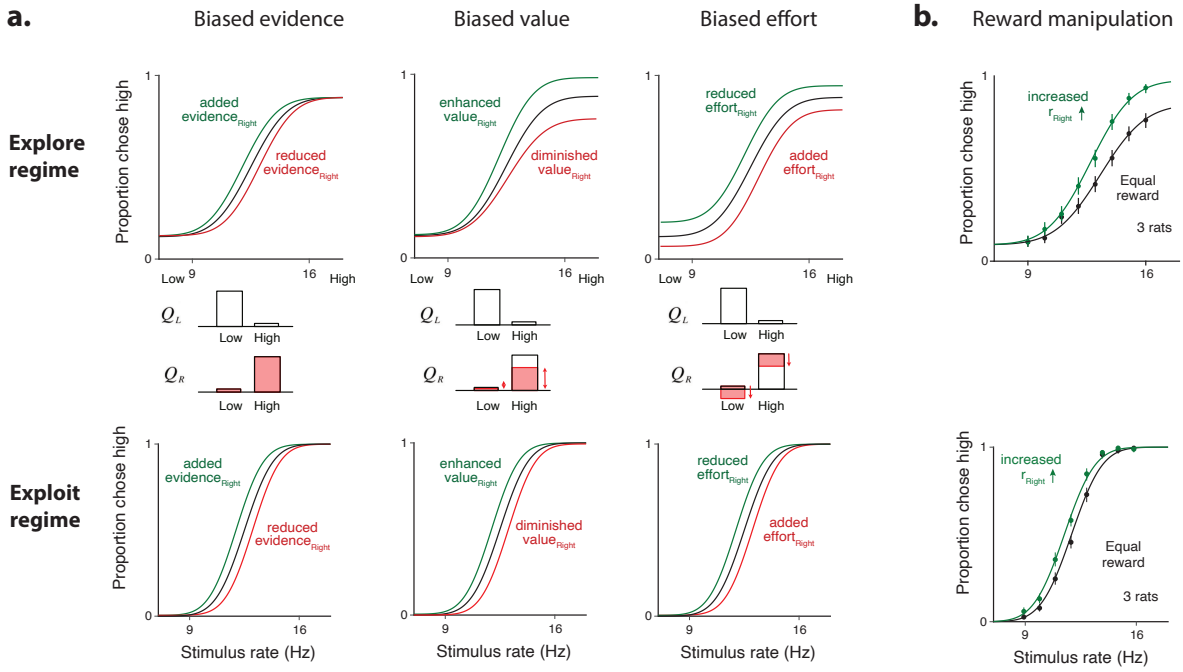


Figure 4 Supplement 2: Psychometric functions with lapses make it possible to assign perturbations effects to specific stages of decision-making (a) (Top row) Model predictions for biased sensory evidence (left), enhanced rightward action value (center) and reduced effort in performing rightward movements (right) in an exploratory regime where lapses are sizeable. The three kinds of perturbations affect decisions at the sensory, value, or motor stages and predict different effects

on lapses. (Middle row) Effects of the three manipulations on the four stimulus-action value pairs. Biasing rightward evidence (left) leaves stimulus-action value pairs unchanged, while biasing the learnt rightward values (center) selectively affects rightward action values on high rates and biasing rightward effort (right) affects both high- and low-rate action values equally. (Bottom row) All three perturbations reduce to the same effect (horizontal shift) in the absence of lapses i.e. in the exploit regime. (b) Example data from 2 rats that experienced the same perturbation: increased rewards on the right port. The rats differ in the extent to which their psychometric functions have lapses. Top: In a psychometric function with lapses, the perturbation (green trace) leads to an interpretable change: the asymmetric change in lapses is only consistent with the explanation that the perturbation enhanced the value of rightward choices (as in (a), top, middle). The perturbation did not drive a change consistent with biased evidence or biased effort. Bottom: In a psychometric function with negligible lapses, the perturbation (red trials) lead to a cryptic change in the psychometric function: the observed shift could equivalently have been driven by biased evidence, value, or effort (as in (a), bottom 3 panels). Therefore, although the perturbation likely caused the same change in the two rats, an experimenter is only able to accurately explain this change in a rat with lapses.

Figure 4 Source data 1: Fit parameters to pooled data across rats

BEHAVIORAL MANIPULATIONS

Multisensory (descriptive: no optimality constraint)

Condition	**No lapse**		**Fixed lapse**				**Restricted lapse**				**Variable lapse**			
	μ	σ	μ	σ	p_{lapse}	$bias_{lapse}$	μ	σ	p_{lapse}	$bias_{lapse}$	μ	σ	p_{lapse}	$bias_{lapse}$
Auditory	12.46	3.43	12.59	3.13	0.06	0.69	12.46	3.43	1E-09	0.05	12.70	2.23	0.25	0.04
Multisensory	12.01	1.87	12.10	1.58	0.06	0.69	11.98	1.82	0.01	0.10	12.13	1.57	0.07	0.03
Visual	12.82	2.94	12.94	2.65	0.06	0.69	12.46	2.56	0.09	0.10	12.54	2.16	0.18	-0.06

Multisensory (theoretical: includes optimality constraint)

	Ideal observer		**Fixed motor error**				**Inattention**				**Exploration**			
	μ	σ	μ	σ	p_{error}	$bias_{error}$	μ	σ	$p_{inattention}$	$bias_{guess}$	μ	σ	β	$bias_{reward}$
Auditory	12.46	3.28	12.61	3.21	0.02	0.03	12.70	2.25	0.24	0.58	12.86	1.85	4.26	0.55
Multisensory	12.00	2.12	12.11	2.06	0.02	0.03	12.14	1.57	0.07	0.75	12.39	1.30	6.93	0.58
Visual	12.81	2.78	12.95	2.70	0.02	0.03	12.54	2.18	0.17	0.34	12.26	1.82	5.16	0.42

Neutral (theoretical: includes optimality constraint)

Auditory	12.82	2.95	13.14	2.62	0.07	0.06	12.83	2.58	0.11	0.49	12.69	1.71	4.56	0.48
Multisensory	12.13	2.19	12.35	1.95	0.07	0.06	12.33	1.44	0.11	0.76	12.57	1.17	6.13	0.60
Neutral	12.29	2.95	12.60	2.62	0.07	0.06	12.83	2.58	0.11	1.00	13.17	1.71	4.56	0.63
Visual	13.02	3.29	13.36	2.92	0.07	0.06	12.62	1.73	0.36	0.39	12.47	1.60	3.55	0.41

Increased reward (Auditory, theoretical)

	μ	σ	μ	σ	p_{error}	$bias_{error}$	μ	σ	$p_{inattention}$	$bias_{guess}$	μ	σ	βr_L	βr_R
Equal reward	13.47	2.86	13.47	2.86	4E-08	0.27	13.95	2.35	0.09	1.00	13.12	1.63	2.33	1.70
increased rR	12.43	2.86	12.43	2.86	4E-08	0.27	12.76	2.35	0.09	0.93	13.12	1.63	2.33	3.74

Decreased reward (Auditory, theoretical)

Equal reward	12.46	3.51	12.91	2.69	0.14	0.82	12.91	2.80	0.11	0.90	13.23	1.85	1.94	3.07
decreased rR	13.24	3.51	13.76	2.69	0.14	0.82	13.88	2.80	0.11	1.00	13.23	1.85	1.94	1.84

Probabilistic reward (Visual, theoretical)

	μ	σ	μ	σ	p_{error}	$bias_{error}$	μ	σ	$p_{inattention}$	$bias_{guess}$	μ	σ	$\beta(r_{LoL} - r_{LoR})$	$\beta(r_{HiR} - r_{HiL})$
p(rHiL) = 0	12.38	2.73	12.00	1.97	0.16	0.27	12.38	2.73	5E-09	0.97	11.79	1.60	3.28	2.25
p(rHiL) = 0.5	12.90	2.73	12.52	1.97	0.16	0.27	12.90	2.73	5E-09	0.31	11.79	1.60	3.28	1.63

NEURAL MANIPULATIONS

	Exploration - biased evidence (kSaline = 0)					**Exploration - biased value (kSaline = 1)**					**Exploration - biased effort (kSaline = 0)**				
	μ	σ	βr_L	βr_R	$k_{muscimol}$	μ	σ	βr_L	βr_R	$k_{muscimol}$	μ	σ	βr_L	βr_R	$k_{muscimol}$
M2 - high rate side inactivation															
Auditory	11.51	2.50	5.49	2.03	-0.90	13.26	1.83	2.74	2.44	0.24	15.00	2.43	2.66	5.88	1.66
Multisensory	12.55	1.68	2.80	4.61	-0.90	12.53	1.25	2.39	4.48	0.24	13.25	1.70	2.95	10.00	1.66
Visual	10.67	2.27	7.44	2.15	-0.90	12.25	1.71	3.22	2.28	0.24	15.00	2.39	2.63	8.96	1.66
M2 - low rate side inactivation															
Auditory	14.00	2.67	2.30	3.59	0.44	13.16	2.15	2.21	2.35	0.51	12.87	2.39	2.32	2.28	-0.76
Multisensory	11.63	1.77	4.38	3.50	0.44	12.10	1.40	3.14	3.71	0.51	11.62	1.57	4.05	3.03	-0.76
Visual	11.87	2.36	4.11	2.51	0.44	11.87	1.85	3.57	2.10	0.51	10.80	2.08	5.94	1.86	-0.76
pStr - high rate side inactivation															
Auditory	12.83	1.97	2.15	2.02	-0.57	12.90	2.08	2.43	2.33	0.49	15.00	3.04	3.18	8.20	1.06
Multisensory	13.28	1.72	1.31	3.90	-0.57	12.60	1.61	1.98	3.67	0.49	12.43	2.37	3.68	5.83	1.06
Visual	4.92	3.49	98.06	1.67	-0.57	12.20	2.55	2.77	1.86	0.49	14.76	3.80	2.97	5.11	1.06
pStr - low rate side inactivation															
Auditory	15.45	3.88	4.50	16.20	0.50	14.36	2.26	1.69	4.46	0.29	11.98	2.86	2.91	2.53	-1.17
Multisensory	15.92	2.92	3.52	28.47	0.50	12.93	1.62	2.24	3.99	0.29	10.50	2.14	7.78	2.64	-1.17
Visual	13.41	4.45	12.22	10.64	0.50	13.24	2.32	2.67	2.12	0.29	10.00	3.24	7.77	1.68	-1.17

Figure 5 - Supplement 1

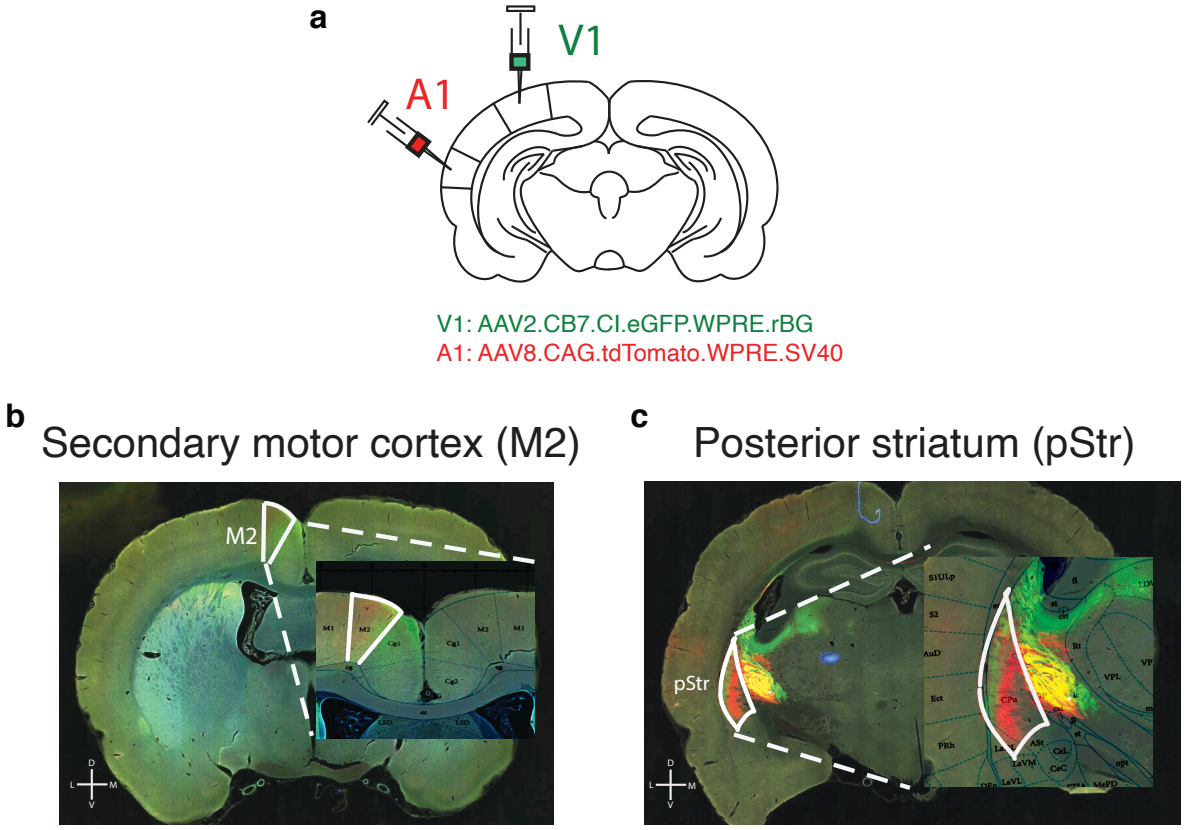
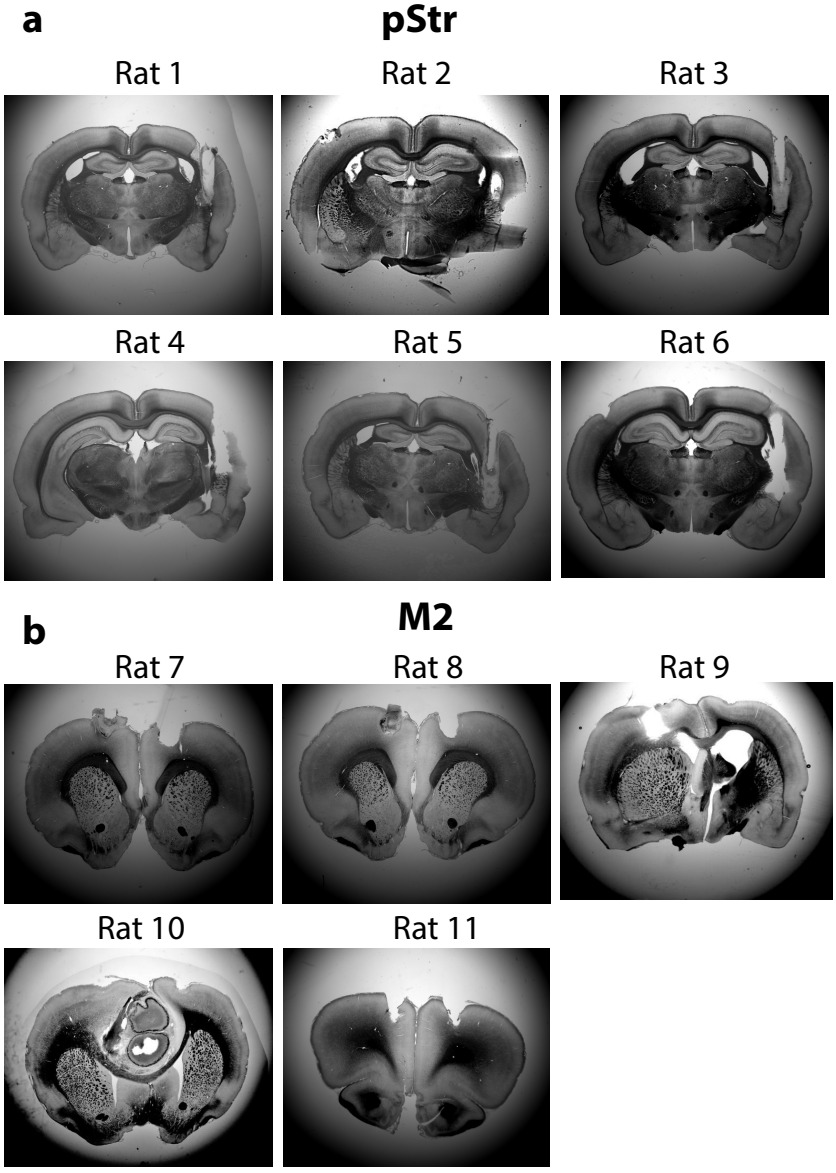


Figure 5 Supplement 1: pStr and M2 receive direct projections from visual and auditory cortex (a) Schematic of tracing experiments. AAV2.CB7.CI.EGFP.WPRE.RBG and AAV2.CAG.tdTomato.WPRE.SV40 constructs were injected unilaterally to primary visual (V1) and auditory (A1) cortices, respectively (V1 coordinates: 6.9 mm posterior to Bregma; 4.2 mm to the right of midline; A1 coordinates: 4.7 mm posterior to Bregma; 7 mm to the right of midline). (b) Secondary motor cortex (M2) receives inputs from V1 and A1 as shown by green and red fluorescence. (c) Posterior striatum (pStr) receives direct inputs from V1 and A1 as shown by green and red fluorescence. Yellow signal medial to pStr reflects overlapping passing fibers.

Figure 5 - Supplement 2



1062

1063 **Figure 5 Supplement 2: Histological slices of implanted rats** Representative coronal slices of all rats

1064 implanted with cannulae for muscimol inactivation experiments. (a) 6 rats were bilaterally implanted in

1065 posterior striatum (pStr). (b) 5 rats were implanted in secondary motor cortex (M2).

1066

Figure 5 - Supplement 3

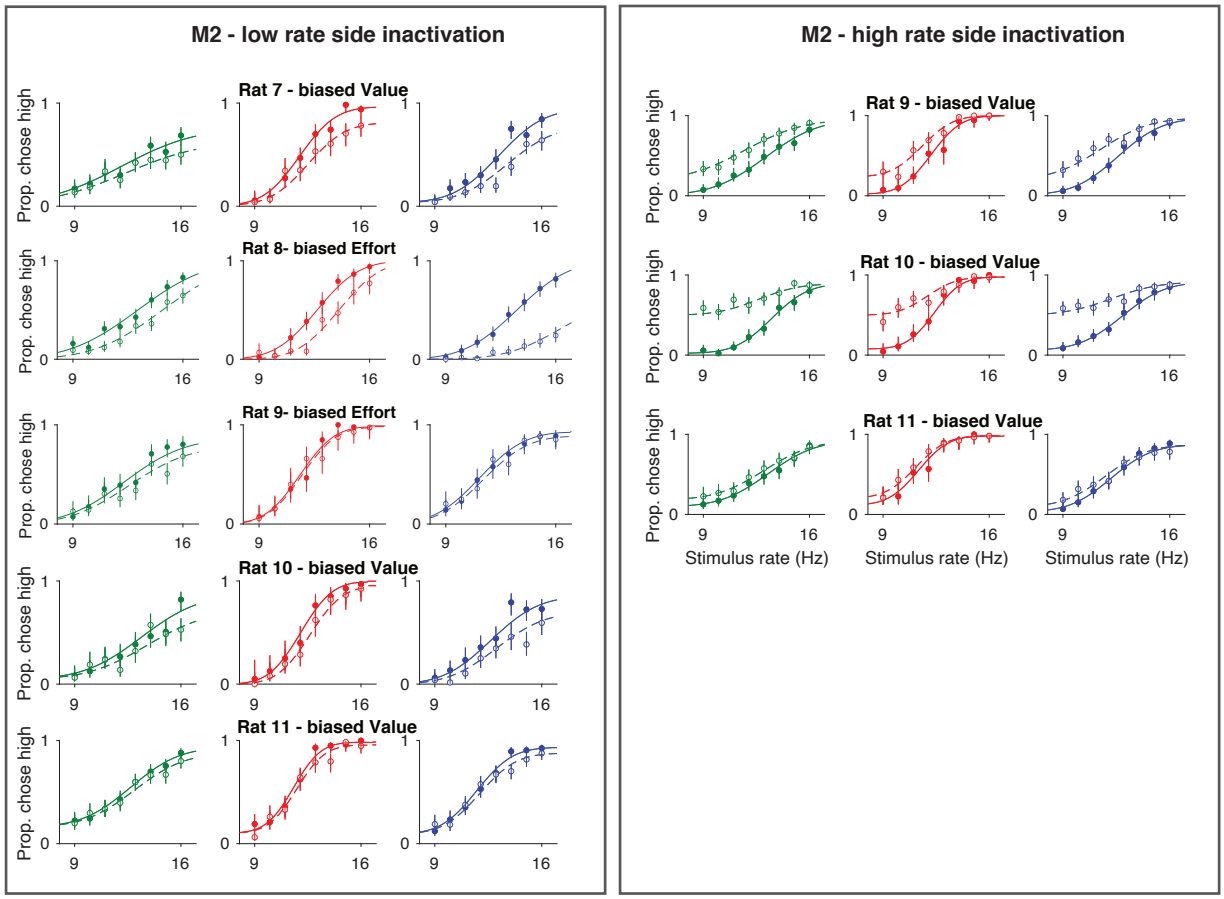


Figure 5 Supplement 3: Single rat performance following M2 inactivation Left: inactivation of the low-rate associated side. Rat shows increased lapses on high-rate trials on all sensory modalities. Right: inactivation of the high-rate associated side. Rat shows increased lapses on low-rate trials on all sensory modalities. Auditory (green), visual (blue) and multisensory (red).

Figure 5 - Supplement 4

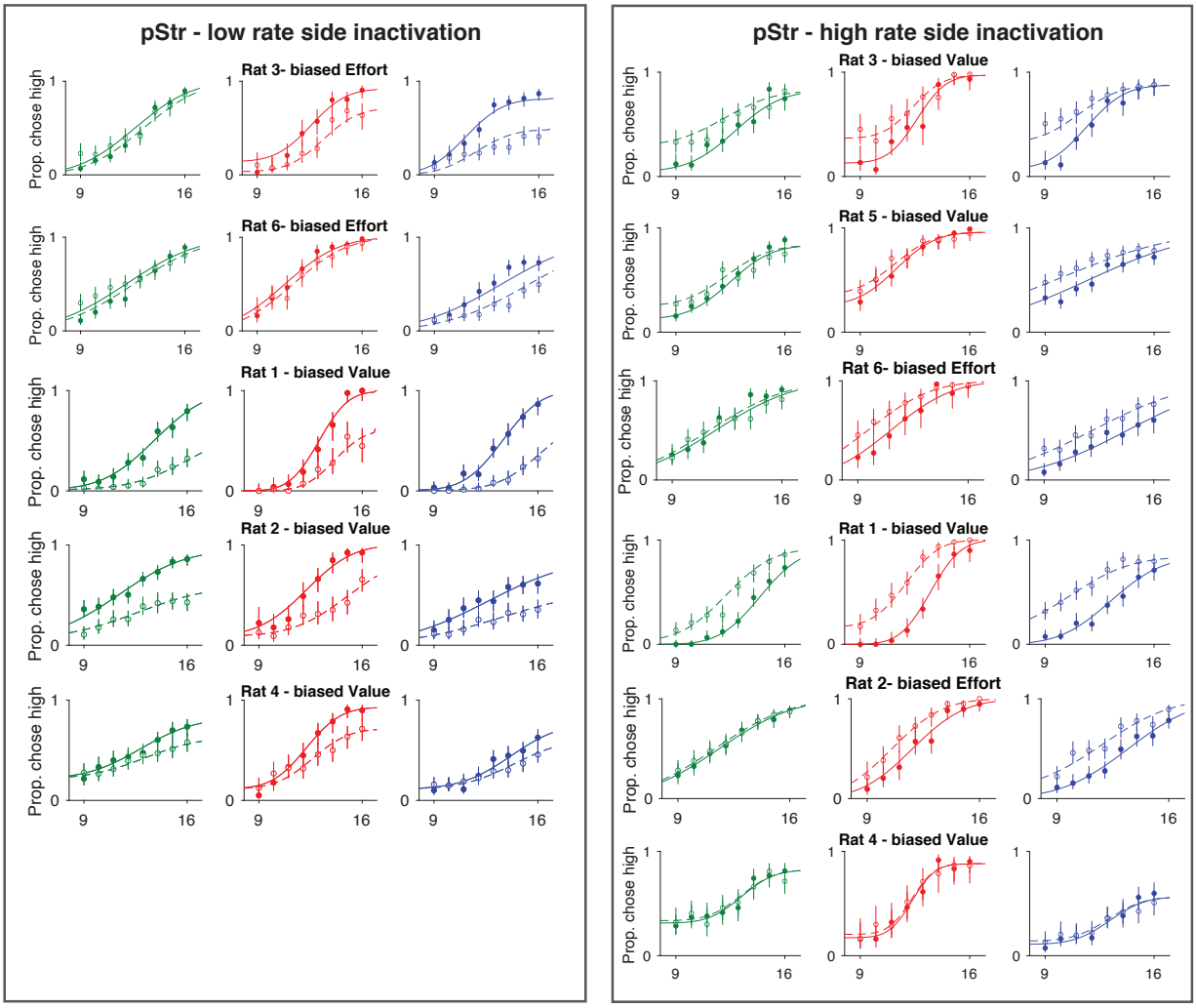


Figure 5 Supplement 4: Single rat performance following pStr inactivation Left: inactivation of the low-rate associated side. Rat shows increased lapses on high-rate trials on all sensory modalities. Right: inactivation of the high-rate associated side. Rat shows increased lapses on low-rate trials on all sensory modalities. Auditory (green), visual (blue) and multisensory (red).

Figure 5 - Supplement 5

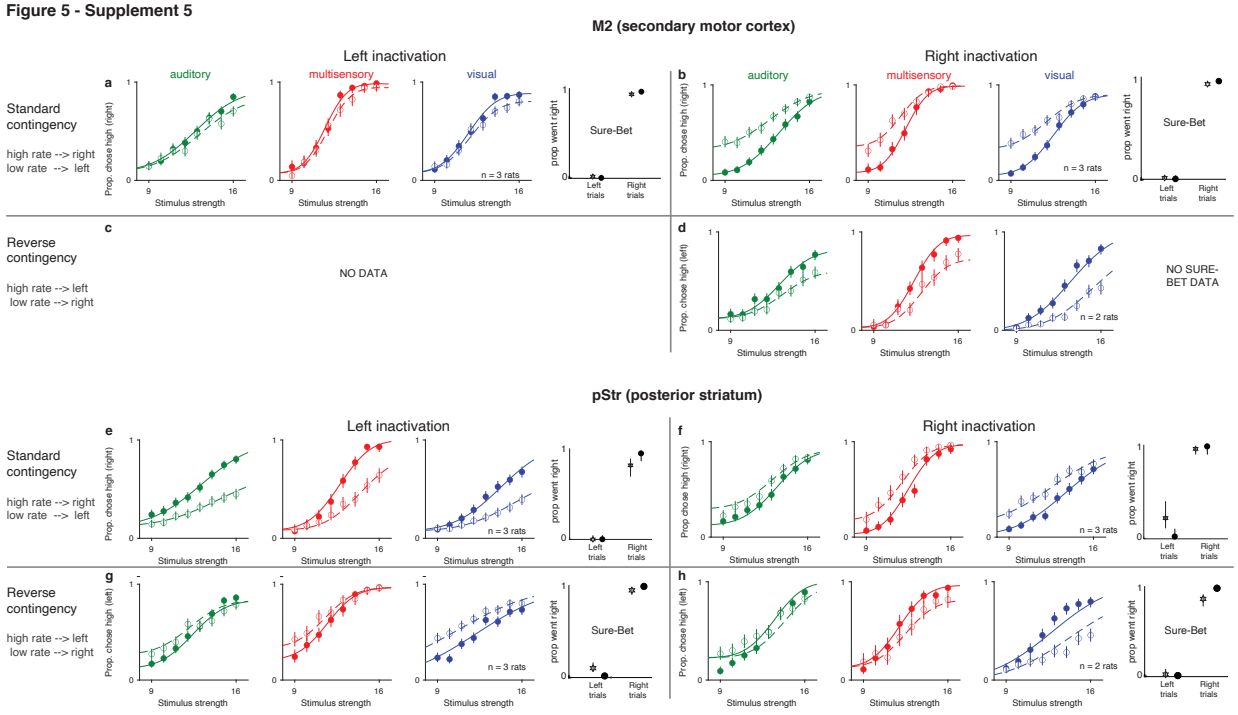


Figure 5 Supplement 5: Unilateral inactivation of M2 or pStr biases performance ipsilaterally and increases contralateral lapses Performance of the same rats shown in Figure 5b depicted as a function of the inactivated side (right or left) and the rate-contingency in which they were trained (standard or reverse), along with fits from the biased value model (Solid lines - Saline, Dotted lines - muscimol). Standard contingency: high rate = go right, low rate = go left; reverse contingency: high rate = go left, low rate = go right. Each quadrant shows 4 plots: 3 psychometrics for rate discrimination trials and one for performance on sure-bet trials. auditory (green), visual (blue) and multisensory (red). (a)-(d) M2 inactivation. (e)-(h) pStr inactivation. (a), (d) Rats trained on the standard contingency and inactivated on the left hemisphere show increased lapses on the high rates (i.e., fewer rightward choices on high rates). No effect on sure-bet trials. (b), (f) Rats trained on the standard contingency and inactivated on the right hemisphere show increased lapses on the low rates (i.e., fewer leftward choices on low rates). No effect on sure-bet trials. (c), (g) Rats trained on the reverse contingency and inactivated on the left hemisphere show increased lapses on the low rates (i.e., fewer

rightward choices on low rates). No effect on sure-bet trials. No data for this condition for M2 inactivation. (d), (h) Rats trained on the reverse contingency and inactivated on the right hemisphere show increased lapses on the high rates (i.e., fewer leftward choices on high rates). No effect on sure-bet trials for pStr inactivated animals; no data for M2 inactivated animals.

Figure 5 - Supplement 6

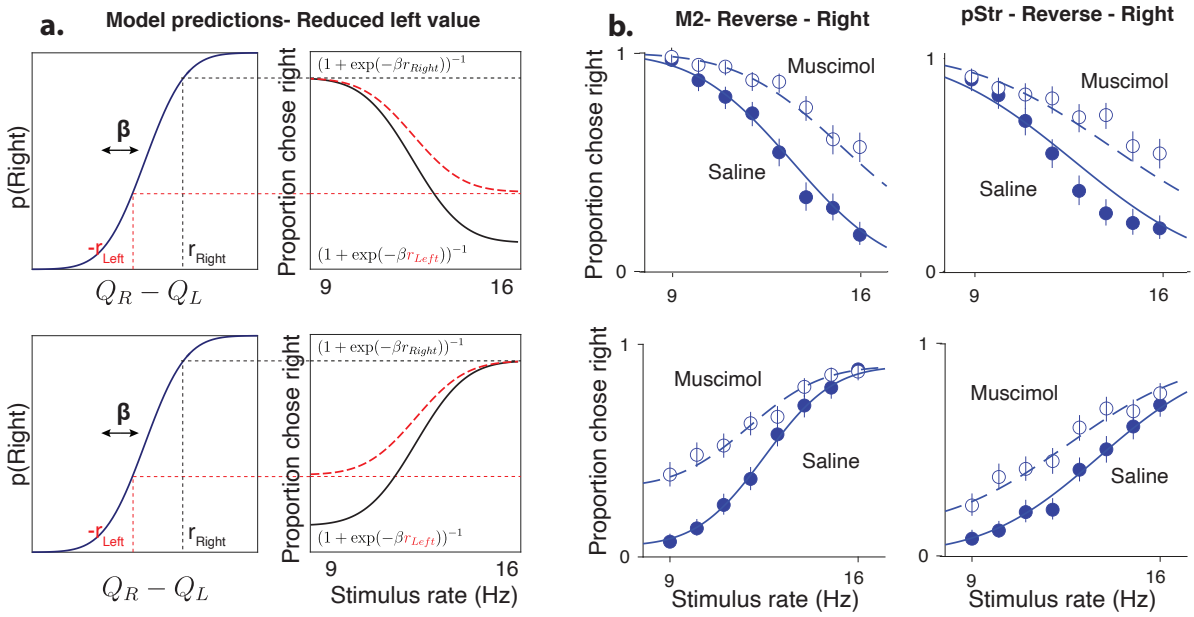


Figure 5 Supplement 6: Inactivations devalue contralateral actions irrespective of associated stimulus (a) Model predictions for rightward inactivations on standard (top) and reversed (bottom) stimulus-response contingencies - in both cases, the model predicts that reduced leftward action values should only affect lapses on the side associated with leftward movements. (b) Inactivation data on visual trials from M2 (left) or pStr (Right) along with fits from the biased value model (Solid lines - Saline, Dotted lines - muscimol) shows a pattern of effects consistent with action value deficits, irrespective of the contingency.

Figure 5 - Supplement 7

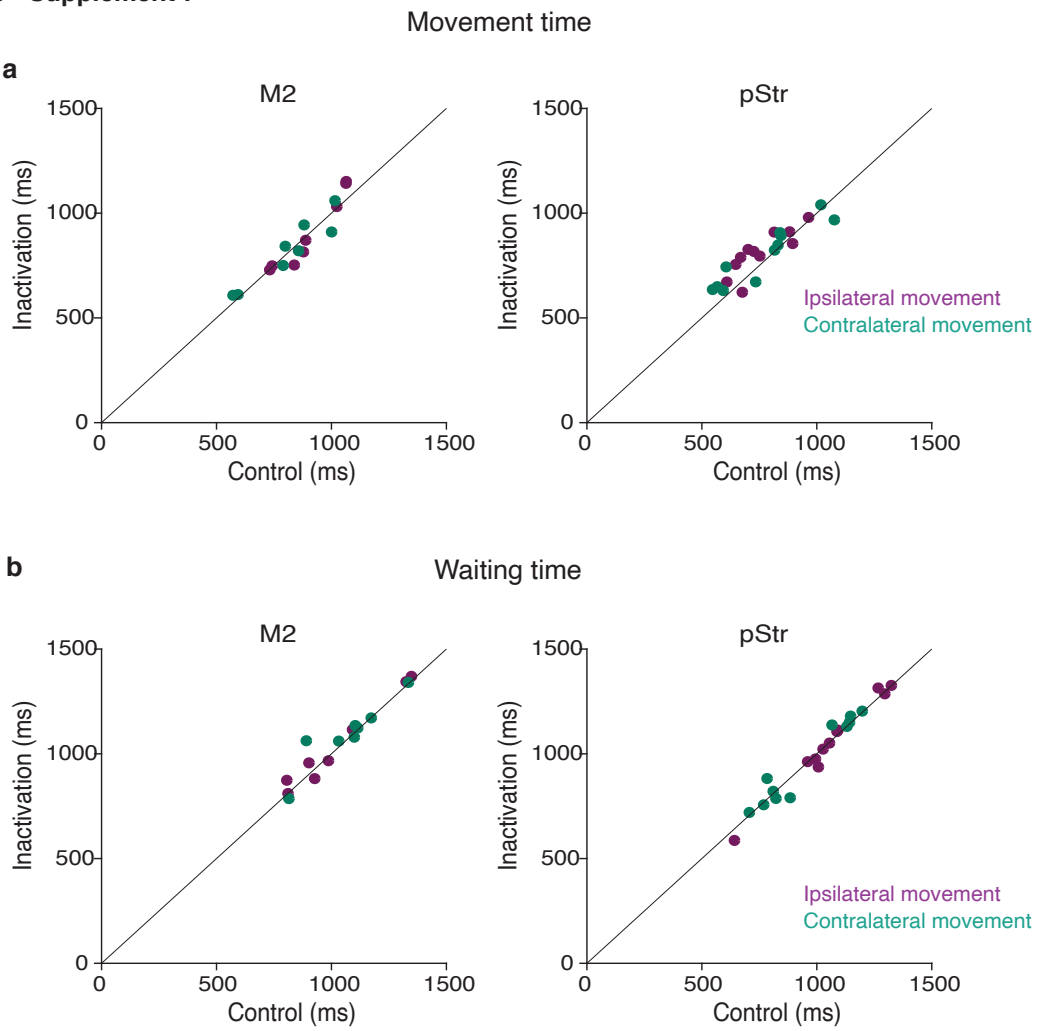


Figure 5 Supplement 7: No significant effect on movement parameters following muscimol inactivation (a) Mean movement times from the center port to the side ports were not significantly different following muscimol inactivation of M2 (left; p = 0.9554 for contralateral, 0.9852 for ipsilateral movements; n=5 rats) or pStr (right; p = 0.6629 for contra, p =0.2615 for ipsi, n=6 rats). Control data on the abscissa is plotted against inactivation data on the ordinate. Purple, movement toward the side ipsilateral to the inactivation site; blue, movement toward the side contralateral to the inactivation site; Error bars (s.e.m.) are not visible because they were obscured by the markers in all cases. (b) Mean wait times in the center port were not significantly

different following muscimol inactivation of M2 (left; p = 0.7612 for contra, p =0.8896 for ipsi, n=5 rats) or pStr (right; p = 0.9128 for contra, p =0.9412 for ipsi, n=6 rats). All p-values were computed from paired t-tests. Error bars (s.e.m.) are not visible because they were obscured by the markers in all cases.

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.

Ashwood, Zoe C, N A Roy, A E Urai, V Aguillon Rodriguez, N Bonacchi, F Cazettes, G A Chapuis, A K Churchland, M Faulkner, F Hu, C Krasniak, I C Laranjeira, G T Meijer, N J Miska, J P Noel, A Pan-vazquez, C Rossant, K Z Socha, I R Stone, M J Wells, C J Wilson, O Winter, IBL Collaboration, and J W Pillow (2019). "State-dependent modeling of psychophysical behavior during decision making". In: *Program No. 241.11. 2019 Neuroscience Meeting Planner. Chicago, IL: Society for Neuroscience.*

Babayan, Benedicte M, Naoshige Uchida, and Samuel J Gershman (2018). "Belief state representation in the dopamine system". In: *Nature communications* 9.1, p. 1891.

Barthas, Florent and Alex C Kwan (2017). "Secondary motor cortex- where sensory meets motor in the rodent frontal cortex". In: *Trends in neurosciences* 40.3, pp. 181–193.

Bays, Paul M, Raquel FG Catalao, and Masud Husain (2009). "The precision of visual working memory is set by allocation of a shared resource". In: *Journal of vision* 9.10, pp. 7–7.

1131 Beeler, Jeff A, Nathaniel D Daw, Cristianne RM Frazier, and Xiaoxi Zhuang (2010). "Tonic
1132 dopamine modulates exploitation of reward learning". In: *Frontiers in behavioral neuroscience*
1133 4, p. 170.
1134 Bertolini, Giovanni, Andrea Wicki, Christian R Baumann, Dominik Straumann, and Antonella
1135 Palla (2015). "Impaired tilt perception in Parkinsons disease- a central vestibular integration
1136 failure". In: *PloS one* 10.4, e0124253.
1137 Bogacz, Rafal, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D Cohen (2006). "The
1138 physics of optimal decision making: a formal analysis of models of performance in two-
1139 alternative forced-choice tasks." In: *Psychological review* 113.4, p. 700.
1140 Busse, Laura, Asli Ayaz, Neel T Dhruv, Steffen Katzner, Aman B Saleem, Marieke L Scholvinck,
1141 Andrew D Zaharia, and Matteo Carandini (2011). "The detection of visual contrast in the
1142 behaving mouse". In: *Journal of Neuroscience* 31.31, pp. 11351–11361.
1143 Carandini, Matteo and Anne K Churchland (2013). "Probing perceptual decisions in rodents". In:
1144 *Nature neuroscience* 16.7, p. 824.
1145 Chakroun, Karima, David Mathar, Antonius Wiehler, Florian Ganzer, and Jan Peters (2019).
1146 "Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making".
1147 In: *BioRxiv*, p. 706176.
1148 Cloherty, Shaun L, Jacob L Yates, Dina Graf, Gregory C DeAngelis, and Jude F Mitchell (2019).
1149 "Motion perception in the common marmoset". In: *bioRxiv*, p. 522888.
1150 Dayan, Peter and Nathaniel D Daw (2008). "Decision theory, reinforcement learning, and the brain".
1151 In: *Cognitive, Affective, & Behavioral Neuroscience* 8.4, pp. 429–453.

Drugowitsch, Jan, Gregory C DeAngelis, Eliana M Klier, Dora E Angelaki, and Alexandre Pouget (2014). "Optimal multisensory decision-making in a reaction-time task". In: *Elife* 3, e03005.

Drugowitsch, Jan and Alexandre Pouget (2018). "Learning optimal decisions with confidence". In: *bioRxiv*, p. 244269.

Drugowitsch, Jan, Valentin Wyart, Anne-Dominique Devauchelle, and Etienne Koechlin (2016). "Computational precision of mental inference as critical source of human choice suboptimality". In: *Neuron* 92.6, pp. 1398–1411.

Ebitz, R Becket, Brianna J Sleezer, Hank P Jedema, Charles W Bradberry, and Benjamin Y Hayden (2019). "Tonic exploration governs both flexibility and lapses". In: *PLoS computational biology* 15.11.

Erlich, Jeffrey C, Bingni W Brunton, Chunyu A Duan, Timothy D Hanks, and Carlos D Brody (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.

Fan, Yunshu, Joshua I Gold, and Long Ding (2018). "Ongoing, rational calibration of reward-driven perceptual biases". In: *Elife* 7, e36018.

Findling, Charles, Vasilisa Skvortsova, Remi Dromnelle, Stefano Palminteri, and Valentin Wyart (2018). "Computational noise in reward-guided learning drives behavioral variability in volatile environments". In: *bioRxiv*, p. 439885.

1172 Flesch, Timo, Jan Balaguer, Ronald Dekker, Hamed Nili, and Christopher Summerfield (2018).
1173 "Comparing continual task learning in minds and machines". In: *Proceedings of the National*
1174 *Academy of Sciences* 115.44, E10313–E10322.
1175 Frank, Michael J, Bradley B Doll, Jen Oas-Terpstra, and Francisco Moreno (2009). "Prefrontal and
1176 striatal dopaminergic genes predict individual differences in exploration and exploitation". In:
1177 *Nature neuroscience* 12.8, p. 1062.
1178 Garrido, Marta I, Raymond J Dolan, and Maneesh Sahani (2011). "Surprise leads to noisier
1179 perceptual decisions". In: *i-Perception* 2.2, pp. 112–120.
1180 Gershman, Samuel J (2015). "A unifying probabilistic view of associative learning". In: *PLoS*
1181 *computational biology* 11.11, e1004567.
1182 — (2018). "Deconstructing the human algorithms for exploration". In: *Cognition* 173, pp. 34–42.
1183 Gold, Joshua I and Long Ding (2013). "How mechanisms of perceptual decision-making affect the
1184 psychometric function". In: *Progress in neurobiology* 103, pp. 98–114.
1185 Green, David M, John A Swets, et al. (1966). *Signal detection theory and psychophysics*. Vol. 1.
1186 Wiley New York.
1187 Guo, Lan, William I Walker, Nicholas D Ponvert, Phoebe L Penix, and Santiago Jaramillo (2018).
1188 "Stable representation of sounds in the posterior striatum during flexible auditory decisions".
1189 In: *Nature communications* 9.1, p. 1534.
1190 Hou, Han, Qihao Zheng, Yuchen Zhao, Alexandre Pouget, and Yong Gu (2018). "Neural correlates
1191 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, Michael Okun, Morgane Moss, Harsha Gurnani, Miles J Wells, Charu Bai Reddy, Kenneth D Harris, and Matteo Carandini (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, L-H Tai, Anthony Zador, and Linda Wilbrecht (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, Tor Lattimore, Laurent Orseau, and Marcus Hutter (2016). "Thompson sampling is asymptotically optimal in general environments". In: *arXiv preprint arXiv:1602.07905*.

Licata, Angela M, Matthew T Kaufman, David Raposo, Michael B Ryan, John P Sheppard, and Anne K Churchland (2017). "Posterior parietal cortex guides visual decisions in rats". In: *Journal of Neuroscience* 37.19, pp. 4954–4966.

Lucas, Christopher G, Sophie Bridgers, Thomas L Griffiths, and Alison Gopnik (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.

Manning, Catherine, Pete R Jones, Tessa M Dekker, and Elizabeth Pellicano (2018). "Psychophysics with children: Investigating the effects of attentional lapses on threshold estimates". In: *Attention, Perception, & Psychophysics*, pp. 1–14.

Mastrogiorgio, Antonio and Enrico Petracca (2018). "Satisficing as an alternative to optimality and suboptimality in perceptual decision making". In: *The Behavioral and brain sciences* 41, e235–e235.

Mendonca, Andre G, Jan Drugowitsch, Maria I Vicente, Eric DeWitt, Alexandre Pouget, and Zachary F Mainen (2018). "The impact of learning on perceptual decisions and its implication for speed-accuracy tradeoffs". In: *bioRxiv*, p. 501858.

Mihali, Andra, Allison Young, Lenard A. Adler, Michael M. Halassa, and Wei Ji Ma (2018). "A Low-Level Perceptual Correlate of Behavioral and Clinical Deficits in ADHD". In: pp. 1–23.

Nikbakht, Nader, Azadeh Tafreshiha, Davide Zoccolan, and Mathew E Diamond (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, Sashank Pisupati, Hien Nguyen, and Anne K Churchland (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, Jeffrey C Erlich, Charles D Kopec, and Carlos D Brody (2017). "Rat prefrontal cortex inactivations during decision making are explained by bistable attractor dynamics". In: *Neural computation* 29.11, pp. 2861–2886.

Pinto, Lucas, Sue A Koay, Ben Engelhard, Alice M Yoon, Ben Deverett, Stephan Y Thiberge, Ilana B Witten, David W Tank, and Carlos D Brody (2018). "An accumulation-of-evidence task using visual pulses for mice navigating in virtual reality". In: *Frontiers in behavioral neuroscience* 12, p. 36.

Pisupati, Sashank, Simon M Musall, Anne E Urai, and Anne K Churchland (2019). "A two stage Bayesian observer predicts the effects of learning on perceptual decisions". In: *Program No. 756.01. 2019 Neuroscience Meeting Planner. Chicago, IL: Society for Neuroscience.*

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, Paul R Schrater, and Anne K Churchland (2012). "Multisensory decision-making in rats and humans". In: *Journal of neuroscience* 32.11, pp. 3726–3735.

Roach, Neil W, Veronica T Edwards, and John H Hogben (2004). "The tale is in the tail: An alternative hypothesis for psychophysical performance variability in dyslexia". In: *Perception* 33.7, pp. 817–830.

Roy, Nicholas G, Ji Hyun Bak, Athena Akrami, Carlos Brody, and Jonathan W Pillow (2018). "Efficient inference for time-varying behavior during learning". In: *Advances in Neural Information Processing Systems*, pp. 5700–5710.

Scott, Benjamin B, Christine M Constantinople, Jeffrey C Erlich, David W Tank, and Carlos D Brody (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.

Siniscalchi, Michael J, Hongli Wang, and Alex C Kwan (2019). "Enhanced population coding for rewarded choices in the medial frontal cortex of the mouse". In: *Cerebral Cortex* 29.10, pp. 4090–4106.

Speekenbrink, Maarten and Emmanouil Konstantinidis (2015). "Uncertainty and exploration in a restless bandit problem". In: *Topics in cognitive science* 7.2, pp. 351–367.

Starkweather, Clara Kwon, Benedicte M Babayan, Naoshige Uchida, and Samuel J Gershman (2017). "Dopamine reward prediction errors reflect hidden-state inference across time". In: *Nature neuroscience* 20.4, p. 581.

Sul, Jung Hoon, Suhyun Jo, Daeyeol Lee, and Min Whan Jung (2011). "Role of rodent secondary motor cortex in value-based action selection". In: *Nature neuroscience* 14.9, p. 1202.

Tai, Lung-Hao, A Moses Lee, Nora Benavidez, Antonello Bonci, and Linda Wilbrecht (2012). "Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value". In: *Nature neuroscience* 15.9, p. 1281.

Wang, Lupeng, Krsna V Rangarajan, Charles R Gerfen, and Richard J Krauzlis (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.

Wilson, Robert C, Andra Geana, John M White, Elliot A Ludvig, and Jonathan D Cohen (2014). "Humans use directed and random exploration to solve the explore–exploit dilemma." In: *Journal of Experimental Psychology: General* 143.6, p. 2074.

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, Timothy D Hanks, Alice Misun Yoon, and Carlos D Brody (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, Nicholas A Steinmetz, Matteo Carandini, and Kenneth D Harris (2019). "Distinct contributions of mouse cortical areas to visual discrimination". In: *bioRxiv*, p. 501627.

Zhou, Baohua, David Hofmann, Itai Pinkoviezky, Samuel J Sober, and Ilya Nemenman (2018). "Chance, long tails, and inference in a non-Gaussian, Bayesian theory of vocal learning in songbirds". In: *Proceedings of the National Academy of Sciences* 115.36, E8538–E8546.

Acknowledgements We thank Matt Kaufman, Simon Musall, Onyekachi Odoemene, Ashley Juavinett, Farzaneh Najafi, Akihiro Funamizu, Priyanka Gupta, Anne Urai, James Roach, Colin Stoneking, Diksha Gupta, Tatiana Engel, Rob Phillips, Tony Zador, Steve Shea and Bo Li for scientific advice and discussions, and Angela Licata, Steven Gluf, Liete Einchorn, Dennis Maharjan, Alexa Pagliaro, Edward Lu and Barry Burbach for technical assistance. We thank Partha Mitra, Alexander Tolpygo and Stephen Savoia for help with slicing and imaging virus injected brains. This work was supported by the Simons Collaboration on the Global Brain, ONR MURI, the Eleanor Schwartz Fund, the Pew Charitable Trust and the Watson School of Biological Sciences.

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

Correspondence Correspondence and requests for materials should be addressed to Anne K. Churchland (email: churchland@cshl.edu).