Article

Integration of sensory evidence and reward expectation in mouse perceptual decision-making task with various sensory uncertainties

Akihiro Funamizu

funamizu@iqb.u-tokyo.ac.jp

Highlights

Tone-frequency discrimination task with two sound durations and biased outcomes

Choices of mice are biased to large reward side in short- than in long-sound trials

Signal detection theory (SDT) suggests a sub-optimal choice behavior

Simple reinforcement learning rule rather than SDT fits the choice behavior

iScience

Article

Integration of sensory evidence and reward expectation in mouse perceptual decision-making task with various sensory uncertainties

Akihiro Funamizu[1,2,3,*]

SUMMARY

In perceptual decision-making, prior knowledge of action outcomes is essential, especially when sensory inputs are insufficient for proper choices. Signal detection theory (SDT) shows that optimal choice bias depends not only on the prior but also the sensory uncertainty; however, it is unclear how animals integrate sensory inputs with various uncertainties and reward expectations to optimize choices. We developed a tone-frequency discrimination task for head-fixed mice in which we randomly presented either a long or short sound stimulus and biased the choice outcomes. The choice was less accurate and more biased toward the large-reward side in short- than in long-stimulus trials. Analysis with SDT found that mice did not use a separate, optimal choice threshold in different sound durations. Instead, mice updated one threshold for short and long stimuli with a simple reinforcement-learning rule. Our task in head-fixed mice helps understanding how the brain integrates sensory inputs and prior.

INTRODUCTION

In a professional tennis tournament, the opponents first serve is typically fast enough that one cannot judge how to return it solely by looking at the ball. When sensory inputs are uncertain, Bayesian theory shows that integration of sensory stimuli with prior knowledge of stimulus probability and action outcomes is essential to optimize behavior (Dayan and Daw, 2008; Gold and Ding, 2013; Körding et al., 2004), as the prediction of serve placement (stimulus probability) and the success rate of forehands and backhands (action outcome) are the keys to returning the first serve. Previous studies have tested the integration of sensory inputs and prior knowledge with a sensory discrimination task by biasing the reward outcomes or stimulus probabilities in humans (Maddox, 2002), monkeys (Feng et al., 2009; Gold and Shadlen, 2007; Hanks et al., 2011; Nomoto et al., 2010; Platt and Glimcher, 1999; Teichert and Ferrera, 2010), pigeons (McCarthy and Davison, 1979; Stüttgen et al., 2011), and rodents (Aguillon-Rodriguez et al., 2021; Lak et al., 2020a; Stoilova et al., 2020). In these studies, subjects tended to select choices associated with a large reward amount or high stimulus probability, indicating that animals integrate the sensory stimulus and the prior.

Signal detection theory (SDT), which is based on Bayesian inference, further predicts that prior-dependent choice biases are also affected by uncertainty in sensory perception (Dayan and Daw, 2008; Gold and Ding, 2013; Maddox, 2002; Stüttgen et al., 2011). When the perceptual uncertainty is low, subjects do not need to use prior knowledge to maximize outcomes in the sensory discrimination task. In contrast, when the uncertainty is high, subjects maximize outcomes by utilizing the prior and bias choices. Previous studies have shown that humans can optimally integrate sensory inputs and priors to make choices (Knill and Pouget, 2004; Körding and Wolpert, 2004). In a perceptual categorization task, humans change their decision threshold depending on the sensory uncertainty and bias the choices (Maddox, 2002). Although animals integrate sensory inputs and priors, it is unclear whether animals change their choice thresholds based on sensory uncertainty to optimize choices.

Here, we developed a single interval forced choice (SIFC) task (Stoilova et al., 2020) for head-fixed mice based on a previous study (Marbach and Zador, 2017), in which mice discriminated a tone frequency to choose between a spout on the left and another on the right. The SIFC task is also known as yes-no task (McCarthy and Davison, 1979; Yeshurun et al., 2008). The head-fixed system in mice will allow us to fully leverage the recent advantages in transgenic, electrophysiology, and optical tools for further experiments

[1]Institute for Quantitative Biosciences, University of Tokyo, Laboratory of Neural Computation, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

[2]Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Meguro-ku, Tokyo 153-8902, Japan

[3]Lead contact

*Correspondence: funamizu@iqb.u-tokyo.ac.jp

([Funamizu et al., 2016](#); [Jun et al., 2017](#); [Rumyantsev et al., 2020](#)). Our task presented either a long or short auditory stimulus in each trial and changed the choice outcome according to the block of trials. Consistent with previous studies in monkeys and rodents ([Brunton et al., 2013](#); [Hanks et al., 2014](#); [Kiani et al., 2008](#)), the long-duration sensory stimulus increased the stimulus evidence; the choice behavior was more accurate with long than with short sound durations. We then found that the choice behavior was less biased by the asymmetric rewards with the long stimuli than with the short stimuli. Analyses with psychometric functions showed that mice had low and high perceptual uncertainty for long and short stimuli, respectively. In both sound durations, the decision threshold to determine left or right choice was similar. This result suggests a suboptimal use of prior knowledge, as stimuli with high perceptual uncertainty require a large shift of threshold in SDT. Instead of the optimal decision threshold, the threshold was shared across sound durations and updated with a simple reinforcement-learning rule in our task. Mice also showed uncertainty-dependent choices when we used high- and low-intensity sounds in the SIFC task. Our task in head-fixed mice helps understanding how the brain integrates the sensory inputs and prior knowledge.

RESULTS

We first introduced our tone frequency discrimination task and showed that mice exploit not only tone frequencies but also tone durations to make more accurate choices in long- than in short-stimulus trials. Psychometric functions showed that the sound durations changed uncertainty in sensory perception, i.e., stimulus sensitivity, while the decision threshold to bias choice was similar. Further analyses with a behavior task and reinforcement learning (RL) model showed that the reward expectations for deciding the choice threshold were shared between the sound durations. The stimulus sensitivity also depended on tone intensities in our SIFC task.

The performance of mice depends on the stimulus duration in a tone-frequency discrimination task

To investigate how different levels of sensory uncertainties affect the choice behavior based on sensory inputs and reward expectations, we developed a tone-frequency discrimination task with two sound durations based on a previous study ([Figure 1](#)A) ([Marbach and Zador, 2017](#)). Mice were placed on a cylindrical treadmill facing two lick spouts. Each trial started by moving the two spouts away from the mice. After a random delay between 1.0 and 2.8 s, a sound stimulus of tone cloud was presented ([Xiong et al., 2015](#); [Zna-menskiy and Zador, 2013](#)). The tone cloud was a mixture of short-duration pure tones (0.03 s) in which the frequency of each tone was either low (5–10 kHz) or high (20–40 kHz). In each trial, the proportion of high-frequency tones was selected from 6 settings (0, 0.25, 0.45, 0.55, 0.75, 1) with low variability in the proportion, and the dominant frequency determined the correct choice. Immediately after the sound ended, the spouts were moved forward, which allowed the mice to choose the left or right spout. The correct or error choice provided 10% sucrose water (2.4 μL) or a noise burst (0.2 s), respectively.

In each trial, we presented either a long (1.0 s) or short (0.2 s) tone cloud, except that the first 40 trials in each session were always long stimuli with 100% low- or high-tone clouds. After the first 40 trials, we biased the reward sizes of the left and right spout for correct choices, with 3.8–1.0 μL or 1.0–3.8 μL (left – right), in each block of 90–120 trials. We defined the asymmetric reward blocks as left or right blocks according to the large-reward side. Left and right blocks were alternated for 4 cycles, and the reward size became 2.4–2.4 μL.

We analyzed the choice behavior of 10 mice, with 71 sessions in total. The high-category tones were associated with the correct choice in the left and right spouts in 5 mice each. In the example session ([Figures 1](#)B and 1C), the mouse succeeded in selecting the correct choices in both the long- and short-stimulus trials. In the short trials, choice behavior was less accurate and more biased toward the large-reward side than in the long trials.

Since mice tended to use the initial part of the stimulus to make choices in previous studies ([Morcos and Harvey, 2016](#); [Odoemene et al., 2018](#)), we first checked whether mice used the whole stimulus duration to make choices ([Figures 1](#)D and S1). In the long-stimulus trials, the choices were correlated with the tone frequency of entire durations (linear mixed-effects model, $p = 1.3 \times 10^{-4} – 5.0 \times 10^{-12}$ (10 mice, 71 sessions)). The choices were also biased by the asymmetric reward conditions ([Figure 1](#)E). The choice performance was better in the long trials than in the short trials, especially when the tone cloud was easy (100% low or high frequency) and moderate (75% low or high) ([Figures 1](#)F and S2). The difference in the correct rate between the long and short trials was also observed when the task had no reward bias ([Figure S3](#)).
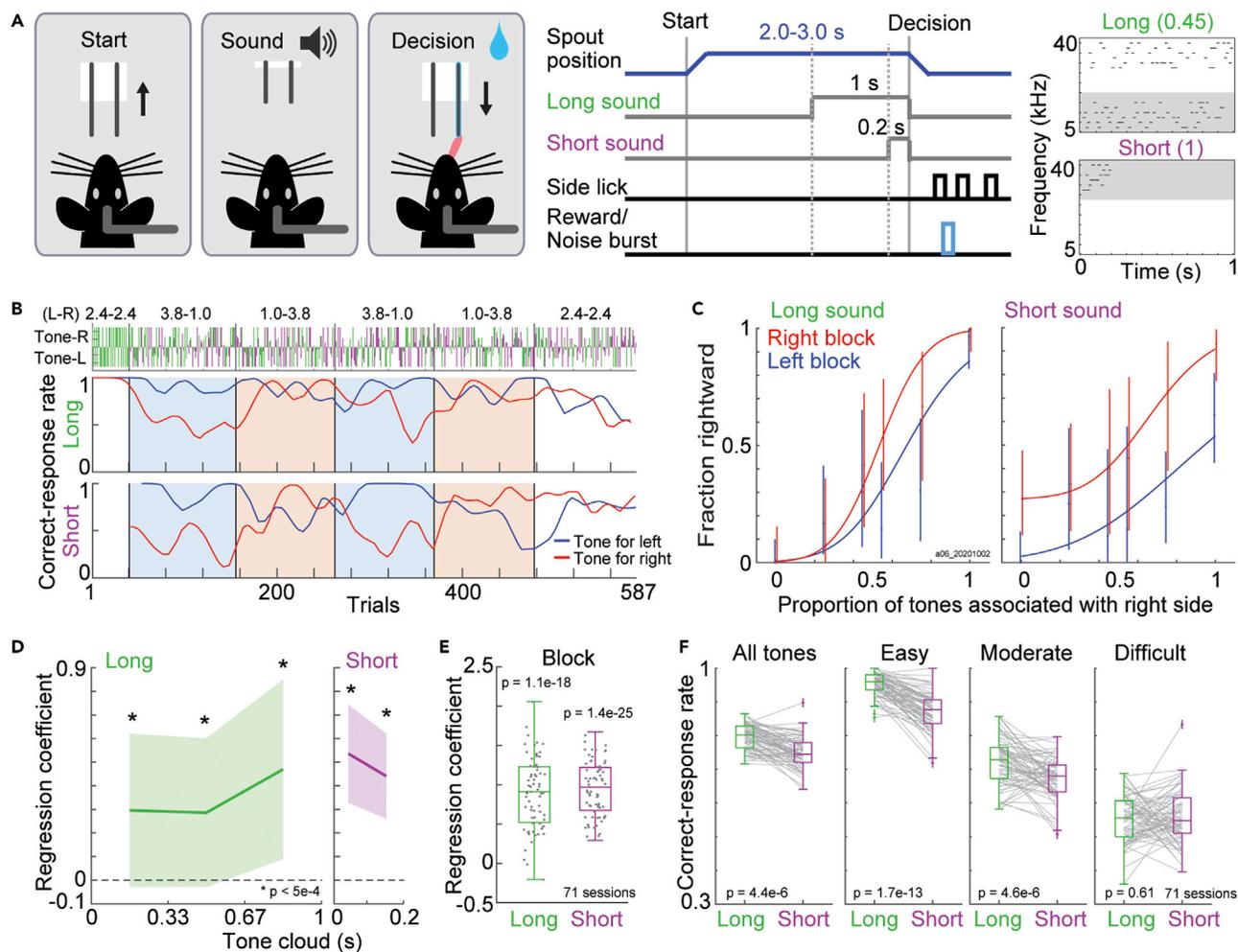
Figure 1. Tone-frequency discrimination task with short and long sound stimuli

(A) Task structure. Each trial started by moving the spouts away from the mouse (Start). After a random delay between 1.0 and 2.8 s, a sound stimulus was presented (Sound). Immediately after the sound ended, the spouts were moved, and the mice were allowed to lick the left or right spout to receive a water reward (Decision). The middle panel shows the time course of the trial. The right panels show example tone clouds with long (top) and short durations (bottom). The parentheses show the proportion of high frequency tones.

(B) Choice behavior in one session. The correct-response rate during tones associated with left or right choice was processed by taking the moving average with a Gaussian function (σ = 10 trials). The top panel shows the stimulus length and proportion of tones associated with left or right choice. Blue and red boxes show the asymmetric-reward blocks (left and right blocks).

(C) Psychometric function of choice behavior in B. Means and 95% confidence intervals.

(D and E) Psychophysical kernels. Logistic regression analyzed how tones at each time point and the asymmetric rewards correlated to the choice behavior. (D, medians and median absolute deviations, linear mixed-effects model, 71 sessions in 10 mice) (E, central mark in box: median; edges of box: 25th and 75th percentiles; whiskers: most extreme data points not considered outliers (beyond 1.5 times the interquartile range), here and throughout). (F) Correct rate of choice behavior for all, easy, moderate, and difficult tone clouds (linear mixed-effects model, 71 sessions in 10 mice).

These results indicate that not only the average tone frequency but also the tone duration affected the choices made by the mice.

Mice have low and high perceptual uncertainty for long and short sound stimuli, respectively

We analyzed the choice behavior of the mice with a psychometric function in which sensory perception was modeled with a truncated Gaussian distribution (one session, Figure 1C; all sessions, Figures 2A and S4) (Masset et al., 2020). The mean of Gaussian was the proportion of tones associated with right-side reward in the tone cloud. The Gaussian distribution was truncated between 0 and 1 (STAR Methods, Equation 2), as the proportion of tones was restricted at the range. Based on the distribution, we computed the cumulative
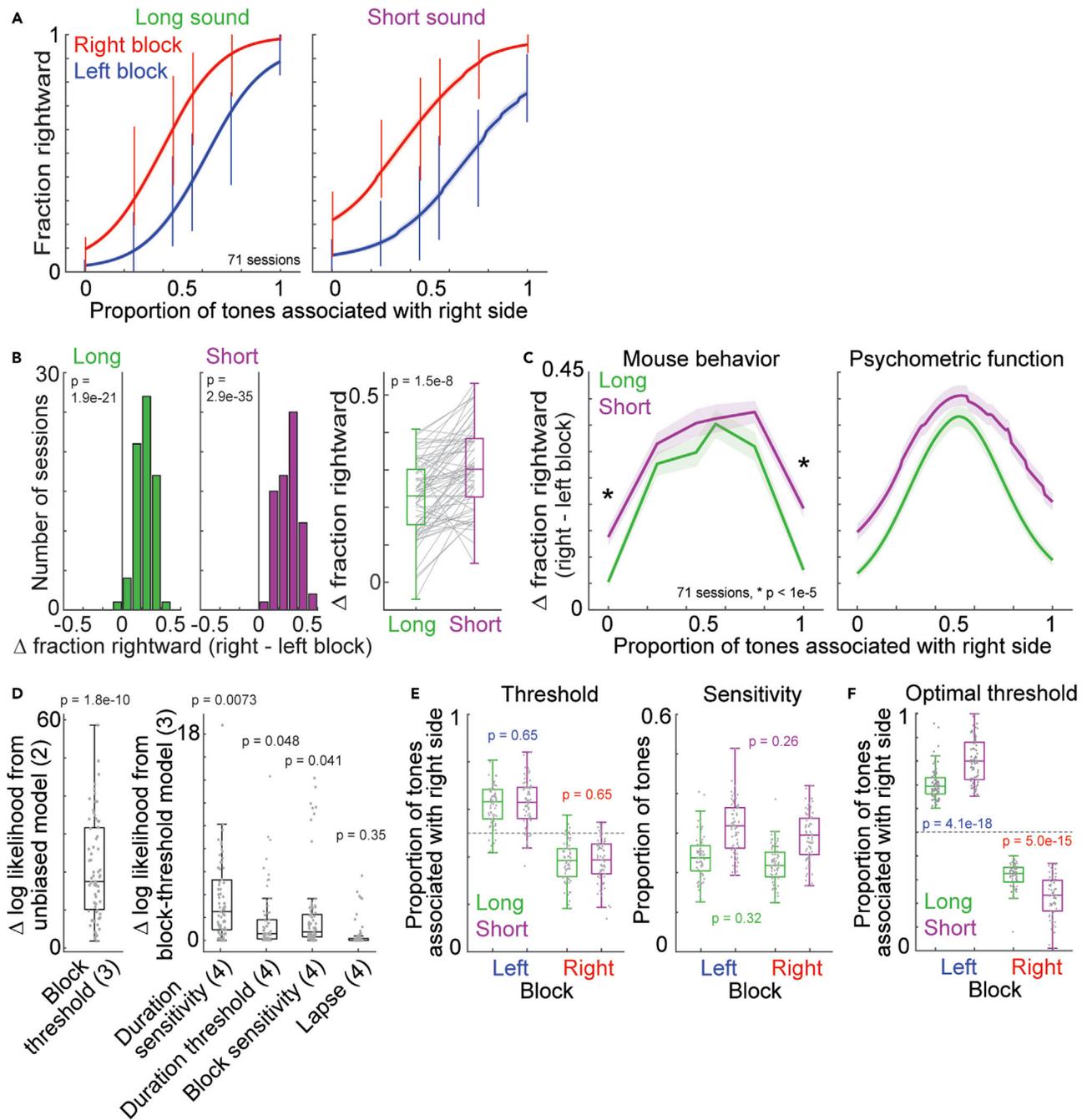
Figure 2. Perceptual uncertainty depends on sound durations

(A) Average psychometric function in each block and duration. Psychometric function had independent decision threshold and stimulus sensitivity in each sound duration and block (4 conditions) to model the behavior of mice. Randomness of choice (lapse rate) was also set in each condition. Error bars show the standard deviation of choice probability (71 sessions).

(B) Block- and duration-dependent choice biases. Δ fraction rightward was the difference in the average rightward choices between the left and right blocks (linear mixed-effects model, 71 sessions in 10 mice).

(C) Comparison of Δ fraction rightward in each tone category. Means and standard errors (linear mixed-effects model, 71 sessions in 10 mice).

(D) Model fitting with likelihood ratio test (71 sessions, 10 mice). Δ log likelihood in the left panel shows how much the log likelihood of an unbiased model, which did not distinguish the blocks and stimulus durations, increased by adding a block-dependent decision threshold (Block threshold). Right panel shows how much the log likelihood of block-threshold model increased by adding each parameter. Parentheses show the number of parameters in the psychometric function.

Figure 2. *Continued*

(E) Decision threshold (threshold) and stimulus sensitivity (sensitivity) in the psychometric function. The decision threshold did not have significant differences between the long and short stimuli, while the stimulus sensitivity was not different between the left and right blocks (linear mixed-effects model, 71 sessions in 10 mice).

(F) Optimal decision threshold. Based on the stimulus sensitivity in (E), we estimated the optimal threshold with signal detection theory. The optimal shift of threshold from the baseline (0.5) was larger in short than in long trials (linear mixed-effects model, 71 sessions in 10 mice).

Gaussian at a decision threshold to estimate the right-choice probability in each trial. By assuming independent decision threshold and stimulus sensitivity (standard deviation of Gaussian) in each tone duration and reward block (2 x 2 = 4 conditions), we investigated the effects of tone durations and reward sizes on the choice behavior. The psychometric function also modeled the randomness of choices (lapse rate), but the following model comparison showed that the lapse rates were zero (Figure 2D), implying that the error choices for easy stimuli (Figures 1C and 2A) were explained by the shift of decision threshold and large stimulus sensitivity for short-duration sounds.

Based on the average rightward choice probability in the psychometric function, we compared the fraction rightward of mice between the left and right blocks and between the long and short sounds (Figure 2B). In both the long- and short-stimulus trials, the choice behavior was biased to the large-reward side (linear mixed-effects model, $p = 1.9 \times 10^{-21}$ and 2.9×10^{-35}). The biases were larger in short- than in long-stimulus trials ($p = 1.5 \times 10^{-8}$). Additionally, the choice bias was particularly large when the tone cloud was 100% high or low tones (linear mixed-effects model, $p = 6.8 \times 10^{-6}$ or 7.4×10^{-12}) (Figure 2C), which was captured by the psychometric function with a truncated Gaussian.

To investigate whether the difference in choices between the long- and short-stimulus trials reflected the optimal integration of sensory stimulus and prior, we investigated the parameters in psychometric function. We first made a simple psychometric function which only had a unique pair of decision threshold and stimulus sensitivity for all the four combinations of tone durations and reward blocks. Adding a parameter in the simple psychometric function always increased the log likelihood; the likelihood ratio test investigated whether the additional parameter significantly improved the fitting (Δ log likelihood in Figure 2D, left) (Daw, 2011; Maddox, 2002). We verified that the psychometric function adding the block-dependent decision threshold (block threshold) significantly improved the fitting than the simple psychometric function (likelihood ratio test, $p = 1.8 \times 10^{-10}$) (Figure 2D, left), indicating that the asymmetric reward blocks biased the choices as reported in previous studies (Daw, 2011; Lak et al., 2020a; Nomoto et al., 2010; Stoilova et al., 2020; Stüttgen et al., 2011). We then found that the psychometric function with block-dependent thresholds and duration-dependent sensitivities fit the choices (likelihood ratio test, $p = 0.0073$), but the parameters for duration-dependent thresholds ($p = 0.048$), block-dependent sensitivities ($p = 0.041$), or lapse rate ($p = 0.35$) did not significantly improve the fitting after Bonferroni correction (Figure 2D, right). These results were consistent with the model fitting with Bayesian information criterion (BIC) (Table S1).

Moreover, we fit the psychometric functions with independent decision thresholds and stimulus sensitivities in each of blocks and sound durations (4 conditions); we found that the decision thresholds were significantly different only by block (linear mixed-effects model, $p = 2.2 \times 10^{-24}$ and 2.2×10^{-24} in long and short trials [10 mice, 71 sessions]), whereas the stimulus sensitivities were different only by stimulus duration ($p = 2.9 \times 10^{-19}$ and 3.4×10^{-17} in left and right blocks) (Figure 2E). These results suggest that the decision thresholds and stimulus sensitivities were shared and independent between the sound durations, respectively. Duration-dependent stimulus sensitivities were also observed in the task without reward bias (Figure S3).

SDT shows that a large amount of sensory uncertainty requires a large choice bias to optimize choice (Dayan and Daw, 2008; Maddox, 2002). Based on the estimated stimulus sensitivities in the psychometric function (Figure 2E, right), we analyzed the optimal decision thresholds for the long and short stimuli. As expected, the optimal choice biases were significantly larger in short- than in long-stimulus trials (linear mixed-effects model, $p = 4.1 \times 10^{-18}$ and 5.0×10^{-15} in left and right blocks [10 mice, 71 sessions]) (Figures 2F and S5). Also, the optimal decision thresholds were significantly larger than the estimated thresholds of mice (linear mixed-effects model, $p = 0.014$–7.4×10^{-8} in all the 4 conditions). These results suggest that although mice had different stimulus sensitivities across sound durations, the choice biases were suboptimal in our task.
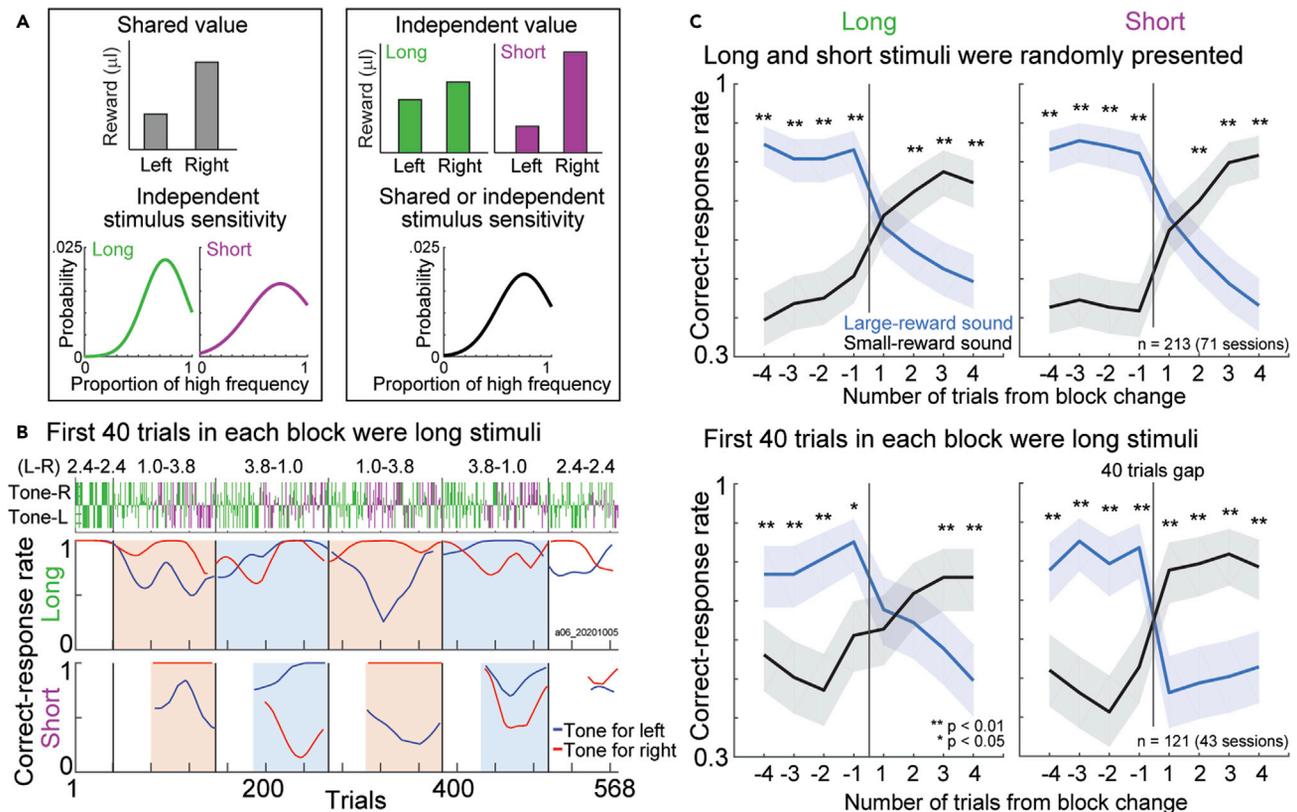
Figure 3. Reward expectations are shared between long and short sounds

(A) Two hypotheses of reward expectations for choices. The shared- and independent-value models hypothesized that the reward expectations for long and short stimuli were jointly or independently updated, respectively.

(B) Example session of a revised task in which the first 40 trials of each block were always long stimuli. The correct-response rate in short trials was flipped from the beginning of each block. Data presentations comply with Figure 1B.

(C) Correct-response rate for moderate and difficult tone clouds around block changes. Blue and black lines show the correct rate of choices associated with large and small rewards in previous blocks, respectively. Means and 95% confidence intervals (linear mixed-effects model; 213 and 121 trials from 71 to 43 sessions, 10 and 9 mice (top and bottom panels)).

The reward expectations of choices are shared between long and short sounds

We proposed two hypotheses to test why the choice biases were suboptimal and became similar between the long- and short-stimulus trials: (i) the reward expectations of left and right choices were shared among different durations; (ii) the reward expectations were independent but learned from similar outcome experiences (Figure 3A). To dissociate the two hypotheses from behavior observations, we made a small change in our task so that the first 40 trials of each block were always long stimuli (Figure 3B). In the revised task, if the reward expectations for long and short stimuli were shared (i), mice could leverage the reward experiences with long stimuli such that the choice in short trials was biased to the large-reward side from the beginning of each block. In contrast, if the reward expectations were independent between the sound durations (ii), the first choice in the short trial was biased to the small-reward side, i.e., the large-reward side in the previous block.

We tested the revised task in 9 mice and analyzed 43 sessions. In the example session (Figure 3B), the choices for short stimuli were biased to the large-reward side from the initial trials of each block. We summarized how the choices in moderate and difficult tone clouds were shifted before and after the block changes (Figure 3C). In our original task in which the long and short stimuli were randomly presented, mice required at least 1 trial to flip the choice biases (comparison of correct rate between the large- and small-reward sides in the first trials of the blocks: linear mixed-effects models, p = 0.54 and 0.48 for long and short trials [213 trials in 71 sessions, 10 mice]). In contrast, in the revised task in which only the long stimuli were presented in the initial trials of each block, mice flipped their choices without direct outcome
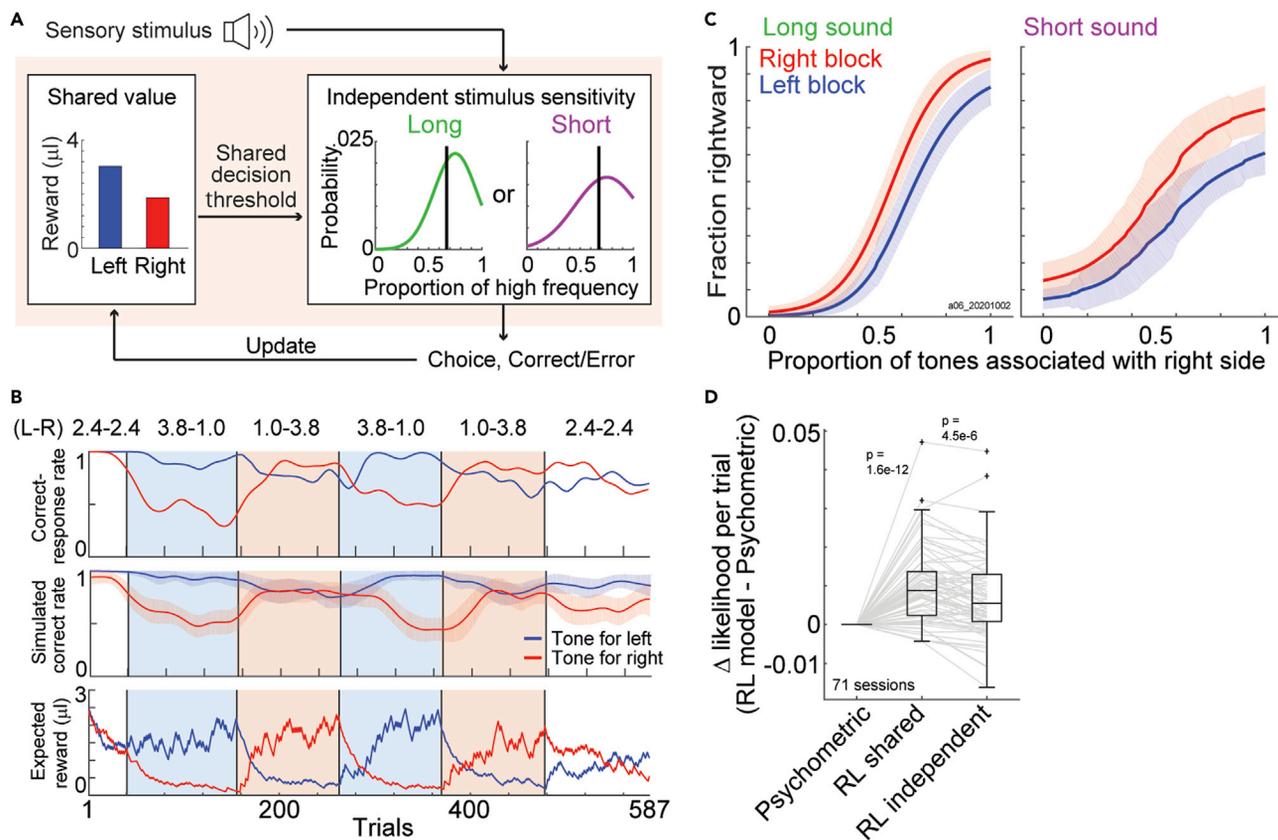
Figure 4. Reinforcement learning model with shared reward expectations and independent stimulus sensitivities for long and short sounds captures choice behavior

(A) Scheme of reinforcement learning (RL) model. The model had shared reward expectations and independent stimulus sensitivities between the long- and short-stimulus trials.

(B) Simulation of choices with RL model in one session. Data are presented in the same manner as in Figure 1B. Based on the fitted parameters in RL model, we simulated the choices of mouse 100 times. Means and standard deviations in simulated correct-response rate.

(C) Simulated psychometric function in B. Means and standard deviations.

(D) Model comparison with cross validation. Our model (RL shared) fit to the choice behavior better than the psychometric function (Psychometric) or an RL model with independent update of reward expectations between the long and short trials (RL independent) (linear mixed-effects model, 71 sessions in 10 mice).

experiences in short trials (p = 0.42 and 2.2 × 10^{-7} [121 trials in 43 sessions, 9 mice]) (direct comparison between the original and revised task for short stimulus: linear mixed-effects models, p = 2.5 × 10^{-4} and 0.0012 for large- and small-reward sounds [213 and 129 trials], respectively). These results indicate that the reward expectations for long and short stimuli were shared between the sound durations.

The RL model shows shared reward expectations and independent stimulus sensitivities for long and short sounds

Trial-by-trial analysis around block changes showed a reward-dependent gradual change in choices (Figure 3C), which was often characterized by a RL model (Funamizu et al., 2012; Ito and Doya, 2015; Stoilova et al., 2020). To understand the computation behind the behavior, we analyzed the choices with RL models (Figure 4A). Our model updated the decision threshold with the outcome experiences of both the long- and short-stimulus trials, while the stimulus sensitivities were independent between the sound durations.

In the example session, our model captured the change in the correct rate by block with the gradual updating of reward expectations (Figure 4B). The RL model also captured (i) the high and low accuracy of choices in long- and short-stimulus trials and (ii) the large effect of asymmetric reward sizes in short trials (Figure 4C). Model comparison with cross validation (STAR Methods) showed that the RL model with shared reward
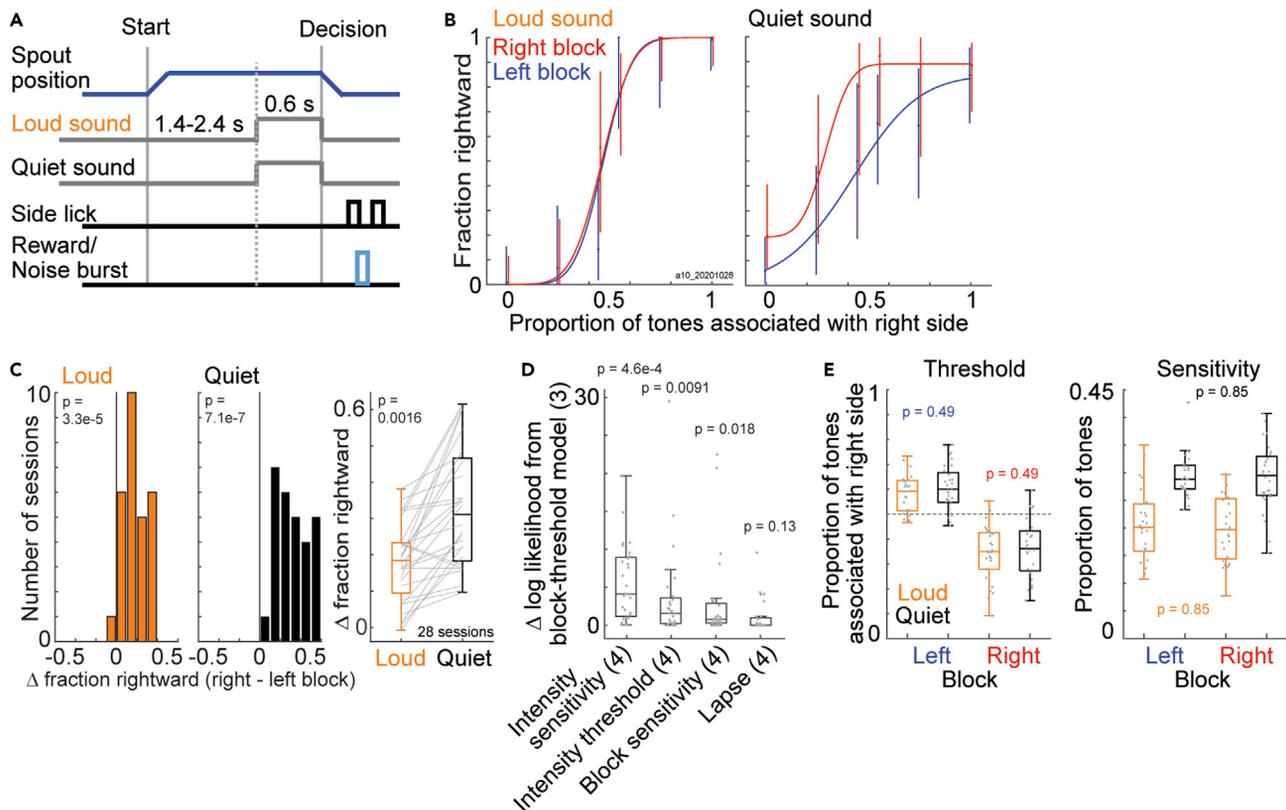
Figure 5. Sound intensity changes perceptual uncertainty in SIFC task

(A) Two-intensity tone-frequency discrimination task. Each trial had either a high- or low-intensity tone cloud with a duration of 0.6 s. Data are presented in the same manner as in Figures 1 and 2.

(B) Psychometric function in one session. Means and 95% confidence intervals.

(C) Comparison of Δ fraction rightward between the high- and low-intensity trials (linear mixed-effects model, 28 sessions in 5 mice).

(D) Model fitting with likelihood ratio test (28 sessions, 5 mice).

(E) Shared decision thresholds and independent stimulus sensitivities between the sound intensities (linear mixed-effects model, 28 sessions in 5 mice).

expectations fit mice behavior better than the psychometric function (linear mixed-effects model, p = 1.6 × 10^{-12}) and an RL model with independent reward expectations between the long and short trials (p = 4.5 × 10^{-6}) (Figure 4D). These results further suggest that mice had different degrees of stimulus sensitivity and shared reward expectations between the sound durations.

Perceptual uncertainty depends on sound intensities in the SIFC task

In addition to the sound durations, we tested how the choice behavior was changed by sound intensities (Meijer et al., 2018). We developed a two-sound intensity tone frequency discrimination task in which either a high- (70 dB SPL [sound pressure level in decibels with respect to 20 μPa]) or low-intensity (30 or 40 dB SPL) tone cloud was presented randomly (Figure 5A). The sound duration was always 0.6 s. The asymmetric reward schedule was identical to the task with two sound durations.

We conducted the two-intensity task in 5 mice and analyzed 28 sessions. The choices were more biased to the large-reward side and less accurate in the low-intensity trials than in the high-intensity trials (example session, Figure 5B; all sessions, Figures 5C and S6 and S7). The analyses with psychometric function showed that the decision thresholds were similar across sound intensities (linear mixed-effects model, p = 0.49 and 0.49 in high- and low-intensity trials), while the stimulus sensitivities were different (p = 3.4 × 10^{-6} and 3.4 × 10^{-6}) (Figures 5D and 5E). These results again suggest the suboptimal choice behavior of mice.

DISCUSSION

Our results showed that mouse behavior in the tone-frequency discrimination task depended not only on the average tone frequency but also on the duration or intensity of the sounds. When we manipulated the sound duration, the choices were more biased to the large-reward side for short stimuli than for long stimuli. Analyses with psychometric functions then showed that the large choice biases were due to the large perceptual uncertainty in short stimuli and were not due to the shift of decision threshold. This result suggests suboptimal behavior in our task: SDT shows that the optimal decision threshold for uncertain stimuli is more biased by asymmetric rewards than for less uncertain stimuli. A revised task and RL model suggested that the reward expectations for deciding decision threshold were shared across sound durations. This result explained the suboptimal behavior in our task.

In SDT, the optimal choice threshold depends on the perceptual uncertainty and reward size of each choice (Figure S5; STAR Methods, Equation 8) (Dayan and Daw, 2008). To compute the optimal threshold, mice needed to understand their own perceptual uncertainty and compute Bayesian inference. Previous studies investigate the optimal integration of sensory inputs and reward expectation in humans (Maddox, 2002) and animals (Feng et al., 2009; Lak et al., 2020a; Nomoto et al., 2010; Pisupati et al., 2021; Stüttgen et al., 2011; Teichert and Ferrera, 2010). In humans, choice biases are often smaller than the optimal (Maddox, 2002). This is consistent with the findings in rodents including our study (Figure 2) (Stoilova et al., 2020). In contrast, monkeys had larger choice biases than the optimal (Feng et al., 2009; Teichert and Ferrera, 2010). Recent study in pigeon found that the reward-dependent choice bias is overshoot at begin and becomes optimal after 1000 trials (Stüttgen et al., 2011). Our study changed the asymmetric reward blocks in about 100 trials, which might not be enough to reach the optimal choice behavior. Another study in human showed that humans changed their decision thresholds depending on sensory uncertainty (Maddox, 2002). This human study changed the sensory uncertainty in every 60 trials, while our study randomly presented the long and short sound stimuli (Figure 1B). Our task schedule possibly affected the optimization of choice behavior in mice.

Instead of Bayes optimal integration of sensory inputs and prior knowledge, a simple way to update the choice threshold is to use recent outcome experiences (Lak et al., 2020b; Rao, 2010). This gradual updating of choice bias was captured by our RL model (Figure 4). Previous studies also found that prior knowledge of stimulus probability and reward amount was updated trial by trial in an RL rule to bias choices (Bell et al., 2016; Lak et al., 2020a; Stoilova et al., 2020).

We developed our auditory discrimination task in a head-fixed system for mice. Head fixation is suitable for large-scale neural recording with two-photon microscopy (Rumyantsev et al., 2020; Stringer et al., 2019) and silicon neural probes such as Neuropixels (Jun et al., 2017; Steinmetz et al., 2019). In addition, the head-fixed system easily controls the parameters of the sensory stimuli. We manipulated sound durations and intensities in our task. We particularly focused on the task with stimulus durations, as the initial phase of stimuli can be identical across durations: mice were required to use the entire stimulus to improve the performance in the long stimuli (Figure 1).

One candidate neural implementation of perceptual uncertainty is probabilistic population coding in which the population of neural activity in sensory cortices encodes the distribution of sensory perception (Knill and Pouget, 2004; Ma et al., 2006). To test population coding, we need to simultaneously record hundreds of neurons. This is suitable for our head-fixed system. Another candidate basis for implementation is the neurons in the orbitofrontal cortex that represent decision confidence (Kepecs et al., 2008; Masset et al., 2020).

In summary, we found that the sound duration or intensity changed the uncertainty in sensory perception and changed the mice behavior in perceptual decision-making task with asymmetric reward conditions. The reward-dependent choice thresholds were determined based on a simple RL rule rather than the optimal Bayesian inference; thus, the choices were suboptimal. Neural recording with two-photon microscopy or high-density silicon probes in our task will provide how the brain represents sensory uncertainty and integrates sensory inputs and prior knowledge to guide behavior.

Limitations of the study

Our study did not investigate the neural correlates of sensory uncertainty which is essential for the choice behavior in our task. Our task randomly presented the long- or short-sound stimuli and switched the reward

amount for correct choices in every 90–120 trials. These task settings might affect the optimality of choice behavior in mice.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHODS DETAILS
  - Tone-frequency discrimination task
  - Data analysis
  - Psychophysical kernels
  - Psychometric function
  - Optimal behavior in the SIFC task
  - Reinforcement learning model
  - Model comparison
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102826.

AUTHOR CONTRIBUTIONS

A.F. designed the experiment, collected and analyzed the data, and wrote the paper.

DECLARATION OF INTERESTS

The author declares no conflict of interests.

REFERENCES

Aguillon-Rodriguez, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., Cazettes, F., Chapuis, G., Churchland, A.K., Dan, Y., Dewitt, E., et al. (2021). Standardized and reproducible measurement of decision-making in mice. Elife 10, e63711.

Bell, A.H., Summerfield, C., Morin, E.L., Malecek, N.J., and Ungerleider, L.G. (2016). Encoding of stimulus probability in macaque inferior temporal cortex. Curr. Biol. 26, 2280–2290.

Brunton, B.W., Botvinick, M.M., and Brody, C.D. (2013). Rats and humans can optimally accumulate evidence for decision-making. Science 340, 95–98.

Constantinople, C.M., Piet, A.T., and Brody, C.D. (2019). An analysis of decision under risk in rats. Curr. Biol.

Daw, N.D. (2011). Trial-by-trial data analysis using computational models. Decis. Making, Affect. Learn. Atten. Perform. XXIII 23, 3–38.

Dayan, P., and Daw, N.D. (2008). Decision theory, reinforcement learning, and the brain. Cogn. Affect. Behav. Neurosci. 8, 429–453.

Feng, S., Holmes, P., Rorie, A., and Newsome, W.T. (2009). Can Monkeys choose optimally when faced with noisy stimuli and unequal rewards? PLoS Comput. Biol. 5, e1000284.

Funamizu, A., Ito, M., Doya, K., Kanzaki, R., and Takahashi, H. (2012). Uncertainty in action-value estimation affects both action choice and learning rate of the choice behaviors of rats. Eur. J. Neurosci.

Funamizu, A., Kuhn, B., and Doya, K. (2016). Neural substrate of dynamic Bayesian inference in the cerebral cortex. Nat. Neurosci.

Gold, J., and Shadlen, M. (2007). The neural basis of decision making. Annu. Rev. Neurosci. 30, 535–574.

Gold, J.I., and Ding, L. (2013). How mechanisms of perceptual decision-making affect the psychometric function. Prog. Neurobiol. 103, 98–114.

Hanks, T., Kiani, R., and Shadlen, M.N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. Elife 3, e02260.

Hanks, T.D., Mazurek, M.E., Kiani, R., Hopp, E., and Shadlen, M.N. (2011). Elapsed decision time

affects the weighting of prior probability in a perceptual decision task. J. Neurosci. *31*, 6339–6352.

Hattori, R., Danskin, B., Babic, Z., Mlynaryk, N., and Komiyama, T. (2019). Area-specificity and plasticity of history-dependent value coding during learning. Cell *177*, 1858–1872.e15.

Ito, M., and Doya, K. (2015). Distinct neural representation in the Dorsolateral, Dorsomedial, and ventral parts of the striatum during fixed- and free-choice tasks. J. Neurosci. *35*, 3499–3514.

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydin, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*, 232–236.

Kepecs, A., Uchida, N., Zariwala, H.A., and Mainen, Z.F. (2008). Neural correlates, computation and behavioural impact of decision confidence. Nature *455*, 227–231.

Kiani, R., Hanks, T.D., and Shadlen, M.N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. J. Neurosci. *28*, 3017–3029.

Knill, D.C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. *27*, 712–719.

Körding, K.P., and Wolpert, D.M. (2004). Bayesian integration in sensorimotor learning. Nature *427*, 244–247.

Körding, K.P., Ku, S.P., and Wolpert, D.M. (2004). Bayesian integration in force estimation. J. Neurophysiol.

Lak, A., Okun, M., Moss, M.M., Gurnani, H., Farrell, K., Wells, M.J., Reddy, C.B., Kepecs, A., Harris, K.D., and Carandini, M. (2020a). Dopaminergic and prefrontal basis of learning from sensory confidence and reward value. Neuron *105*, 700–711.e6.

Lak, A., Hueske, E., Hirokawa, J., Masset, P., Ott, T., Urai, A.E., Donner, T.H., Carandini, M., Tonegawa, S., Uchida, N., et al. (2020b). Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. Elife *9*, e49834.

Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. Nat. Neurosci.

Maddox, W.T. (2002). Toward a unified theory of decision criterion learning in perceptual categorization. J. Exp. Anal. Behav. *78*, 567–595.

Marbach, F., and Zador, A.M. (2017). A self-initiated two-alternative forced choice paradigm for head-fixed mice. bioRxiv. https://doi.org/10.1101/073783.

Masset, P., Ott, T., Lak, A., Hirokawa, J., and Kepecs, A. (2020). Behavior- and modality-general representation of confidence in orbitofrontal cortex. Cell *182*, 112–126.e18.

McCarthy, D., and Davison, M. (1979). Signal probability, reinforcement and signal detection. J. Exp. Anal. Behav. *32*, 373–386.

Meijer, G.T., Pie, J.L., Dolman, T.L., Pennartz, C.M.A., and Lansink, C.S. (2018). Audiovisual integration enhances stimulus detection performance in mice. Front. Behav. Neurosci. *12*, 1–12.

Morcos, A.S., and Harvey, C.D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. Nat. Neurosci.

Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nat. Neurosci.

Nomoto, K., Schultz, W., Watanabe, T., and Sakagami, M. (2010). Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. J. Neurosci. *30*, 10692–10702.

Odoemene, O., Pisupati, S., Nguyen, H., and Churchland, A.K. (2018). Visual evidence accumulation guides decision-making in unrestrained mice. J. Neurosci. *38* (47), 10143–10155. https://doi.org/10.1523/JNEUROSCI.3478-17.2018.

Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., and Churchland, A.K. (2021). Lapses in perceptual judgments reflect exploration. eLife *10* (e55490). https://doi.org/10.7554/eLife.55490.

Platt, M.L., and Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex.

Nature *400* (6741), 233–238. https://doi.org/10.1038/22268.

Rao, R.P.N. (2010). Decision making under uncertainty: a neural model based on partially observable Markov decision processes. Front. Comput. Neurosci. *4* (146). https://doi.org/10.3389/fncom.2010.00146.

Rumyantsev, O.I., Lecoq, J.A., Hernandez, O., Zhang, Y., Savall, J., Chrapkiewicz, R., Li, J., Zeng, H., Ganguli, S., and Schnitzer, M.J. (2020). Fundamental bounds on the fidelity of sensory cortical coding. Nature *182*, 112–126.e18.

Steinmetz, N.A., Zatka-Haas, P., Carandini, M., and Harris, K.D. (2019). Distributed coding of choice, action and engagement across the mouse brain. Nature *576*, 266–273.

Stoilova, V.V., Knauer, B., Berg, S., Rieber, E., Jäkel, F., and Stüttgen, M.C. (2020). Auditory cortex reflects goal-directed movement but is not necessary for behavioral adaptation in sound-cued reward tracking. J. Neurophysiol. *124*, 1056–1071.

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K.D. (2019). High-dimensional geometry of population responses in visual cortex. Nature *571*, 361–365.

Stüttgen, M.C., Yildiz, A., and Güntürkün, O. (2011). Adaptive criterion setting in perceptual decision making. J. Exp. Anal. Behav. *96*, 155–176.

Teichert, T., and Ferrera, V.P. (2010). Suboptimal integration of reward magnitude and prior reward likelihood in categorical decisions by monkeys. Front. Neurosci. *4*, 1–13.

Xiong, Q., Znamenskiy, P., and Zador, A.M. (2015). Selective corticostriatal plasticity during acquisition of an auditory discrimination task. Nature *521*, 348–351.

Yeshurun, Y., Carrasco, M., and Maloney, L.T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. Vis. Res *48*, 1837–1851.

Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. Nature *497*, 482–485.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Mouse behavioral data	This paper; Mendeley Data	https://dx.doi.org/10.17632/jp9ytffdt9.1
Experimental models: Organisms/strains		
CBA/J Mice	Charles River Japan	JAX stock 000656
Software and algorithms		
Matlab 2016b	Mathworks	2016b
Reinforcement learning model	This paper; Zenodo	https://dx.doi.org/10.5281/zenodo.5054897
Other		
Bpod framework (control for behavioral task)	Sanworks	r0.5
Speaker	Avisoft Bioacoustics	60108
Microphone for sound calibration	Brüel and Kjaer	Type 4939

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to Akihiro Funamizu (funamizu@iqb.u-tokyo.ac.jp).

Materials availability

Not applicable.

Data and code availability

Behavioral datasets have been deposited at Mendeley and are publicly available as of the date of publication. Original code has been deposited at Zenodo and is publicly available as of the date of publication. The DOIs for datasets and code are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We used 11 male CBA/J mice (Charles River, Japan), 8 to 9 weeks of age at the start of behavioral training. All animal procedures were approved by the Animal Care and Use Committee at the Institute for Quantitative Biosciences (IQB), University of Tokyo. Mice were housed in a temperature-controlled room with a 12 h/12 h light/dark cycle. All experiments were done during the dark cycle.

The mouse surgery was based on the procedure used at the previous institute of the first author (Cold Spring Harbor Laboratory). Before surgery, mice were restricted to 1.5 mL of water per day for at least two weeks. The weight of each mouse was checked daily to avoid dehydration. Two days before surgery, mice received free access to water. For a behavior task with head-restrained mice, we implanted a custom-designed lightweight head bar. Mice were anesthetized with isoflurane (1.5% at induction, below 1% at maintenance) with an additional analgesic (meloxicam 2 mg/kg, subcutaneous) and eye ointment. The mice were placed in a stereotaxic apparatus. The scalp was removed above the entire cortical area. The skull was cleaned with povidone iodine and hydrogen peroxide. The head bar was attached to the skull with superbond adhesive (Sun Medical or Parkell S380) and cyanoacrylate glue (Zap-A-Gap, PT03) (Musall et al., 2019). After surgery, the mice received free access to water until they recovered.

METHODS DETAILS

Tone-frequency discrimination task

After recovery from surgery, behavioral training started. The weight of each mouse was carefully monitored, and additional water was given after daily training to avoid dehydration. The behavioral setup

and the task have been described previously (Marbach and Zador, 2017). All the training was performed inside a custom sound booth. Mice were head-fixed and positioned over a cylinder treadmill. One speaker (#60108, Avisoft Bioacoustics) was placed diagonally to the right of the mice for auditory stimulation. The speaker was calibrated with a free-field microphone (Type 4939, Brüel and Kjaer). Water was delivered through two spouts connected to a custom lick detection circuit. The spouts were moved back and forth in each trial, and their position was controlled by an Arduino with a servo motor (hsb-9475sh, Hitech). Water was calibrated before experiments. The behavioral system was controlled by a custom MATLAB (MathWorks) program running on the Bpod framework (https://sanworks.io) on Windows.

The task required mice to select the left or right spout depending on the frequency of the sound stimulus (Figure 1). Each trial started by retracting the two spouts away from mice. After a random interval between 1.0 and 2.8 s, a sound stimulus was started. The sound stimulus was a tone cloud consisting of a series of 30 ms pure tones with rise/decay ramps of 3 ms at a rate of 100 tones per second (98 and 18 tones for long (1.0 s) and short (0.2 s) stimuli) (Xiong et al., 2015; Znamenskiy and Zador, 2013). The frequency of each tone was sampled from 18 logarithmically spaced slots (5 to 40 kHz). The tone cloud in each trial contained the low (5 – 10 kHz) and high frequency tones (20 – 40 kHz) and was categorized as low or high depending on the dominant frequency. The proportion of high tones in each tone cloud was selected from 6 settings (0, 0.25, 0.45, 0.55, 0.75, 1) with a probability of (25%, 12.5%, 12.5%, 12.5%, 12.5%, 25%, i.e., 2:1:1:1:1:2). In each trial, the proportion of high tones had low variability from the settings except for the 100% low or high tone.

Immediately after the sound end, the two spouts were moved forward, and mice could select the left or right spout. For each mouse, the high- or low-category tone was associated with a 10% sucrose water reward in either the left or right spout as a correct trial. The selection of the opposite spout triggered a noise burst (0.2 s) marking an error trial. The outcome was delivered immediately after the choice. When mice did not select a spout within 30 s of the start of the trial, a new trial started.

In the two-sound-duration tone frequency discrimination task (Figure 1), a duration of either 1.0 or 0.2 s was randomly selected in each trial (long or short trial), except that the first 40 trials in each session were always long stimuli with 100% low- or high-tone clouds. The intensity of the tone cloud was constant in each trial but sampled from either 60, 65 or 70 dB SPL (sound pressure level in decibels with respect to 20 μPa) to prevent mice from using loudness for tone discrimination. After the first 40 trials, the reward amount of the left and right spout was switched by a block of 90 – 120 trials. In each block, the reward size for a correct left or right choice was either 1.0 and 3.8 μl, respectively, or 3.8 and 1.0 μl, respectively. We defined the left and right blocks depending on the large reward for the corresponding choice. After mice had experienced asymmetric reward blocks 4 times, the reward sizes of the left and right spouts became identical (2.4 – 2.4 μl). If mice did not complete the 4 blocks, we did not use the session for analysis.

In the two-sound-intensity tone-frequency discrimination task (Figure 5), the sound duration was always 0.6 s, while the sound intensities were randomly sampled from low (30 or 40 dB SPL) or high (70 dB SPL) in each trial. In one mouse, we tested different combinations of sound intensities (combination of 30-70 dB SPL or 40-70 dB SPL) and used the sessions with 30-70 dB SPL for analyses, as the combination had a larger intensity difference. The first 40 trials in each session were always high-intensity sounds with 100% low- or high-tone clouds. The reward schedule was identical to that of the two-duration task.

Data analysis

All analyses were performed with MATLAB (MathWorks). In each session, trials in which mice succeeded in selecting the left or right spout were analyzed. In the two-duration task, we analyzed 71 sessions from 10 out of 11 mice (mouse a01 – a10: 4, 9, 8, 8, 10, 6, 6, 8, 6, 6 sessions) (Figure S1). In one mouse, we obtained data from only one session, which was not included in the analyses. Additionally, we did not use the session for analyses when mice did not select correct choices at least 80% of the time for the 100% low- and high-tone clouds. In the two-duration task without reward bias (Figure S2), we analyzed 37 sessions from 9 mice (mouse a02 – a10: 5, 4, 5, 6, 3, 4, 4, 3, 3 sessions). In the two-intensity task (Figure S4), we analyzed 28 sessions from 5 mice (mouse a04, a05, a06, a08, a10: 6, 6, 6, 6, 4 sessions).

Psychophysical kernels

We used logistic regression (MATLAB: glmfit) to quantify which tone-cloud timing and asymmetric reward blocks were correlated with the choice in each trial t:

$$\log\left(\frac{p(right, t)}{1 - p(right, t)}\right) = \beta_0 + \beta_1 b(t) + \sum_{time=0}^{T} \beta_{time} E_{right}(time, t), \qquad \text{(Equation 1)}$$

where $p(right, t)$ was the rightward choice probability at trial t. β was a regression coefficient. $E_{right}(time, t)$ was the proportion of tone frequency associated with rightward choice in a tone cloud at trial t, time window $time$. $b(t)$ was the asymmetric reward block at trial t (left or right). $E_{right}(time, t)$ and $b(t)$ were normalized. For the psychophysical kernels, we used only the trials with moderate (75% low or high frequency) and difficult (55% low or high) tone clouds.

Psychometric function

We assumed that when a tone cloud of $E_{right}(t)$ was presented at trial t, its noisy sensory perception $\widehat{E}$ had a Gaussian distribution with the mean $E_{right}(t)$ truncated between 0 and 1 of stimulus x, as the proportion of tone frequency in the sound stimulus was bounded at the range:

$$\widehat{E} = N(x|E_{right}(t), \sigma^2) \Big/ \int_0^1 N(x|E_{right}(t), \sigma^2) dx = ZN(x|E_{right}(t), \sigma^2), \qquad \text{(Equation 2)}$$

where σ was the perceptual uncertainty (standard deviation). Z truncated the Gaussian distribution between 0 and 1 here and throughout. Based on the noisy sensory perception, we estimated the rightward choice probability by analyzing the cumulative Gaussian:

$$p(right, t) = \lambda_1 + (1 - \lambda_1 - \lambda_2) \int_A^1 ZN(x|E_{right}(t), \sigma^2) dx,$$
$$A = \beta_0 + \beta_1 b(t) + \beta_2 s(t),$$
$$\sigma^2 = \beta_3 + \beta_4 b(t) + \beta_5 s(t), \qquad \text{(Equation 3)}$$

where $b(t)$ and $s(t)$ were -1 or 1 for the left or right block (b) and the short or long stimulus (s), respectively. β_{0-5} were regression coefficients. β_{0-2} determined the threshold for selecting the left or right choice. β_1 and β_2 quantified the choice bias by the blocks and sound durations, respectively. β_{3-5} determined the slope of the psychometric curve of behavior (stimulus sensitivity or perceptual uncertainty). $\lambda_{1,2}$ were lapse rates that quantified the randomness of choices in all stimuli. We could set different lapse rates in each block and sound duration, but the following model fitting showed that the lapse rate was zero in our task (Figure 2D, Model comparison). This implies that the error choices for easy stimuli (Figures 1C and 2A) were explained by the shift of decision threshold and large stimulus sensitivity for short-duration sounds. In the two-intensity task (Figure 5), we used Equation 3 to analyze the psychometric function except that $s(t)$ was the stimulus intensity (-1 or 1 for low or high intensity).

Optimal behavior in the SIFC task

SDT derived the optimal behavior in our SIFC task by estimating the category C (low or high) of each trial from the tone cloud. The proportion of high frequency in the tone cloud fluctuated trial by trial; we thus assumed that the proportion of high frequency in the low- and high-category tones had a uniform distribution of stimulus x from 0 to 0.5 and from 0.5 to 1, respectively. The likelihood of a sensory stimulus in category C_i, $P(x|C_i)$, was defined as follows:

$$P(x|C_i) = \int_0^1 P(E_j|C_i) ZN(x|E_j, \sigma^2) dE_j. \qquad \text{(Equation 4)}$$

$P(E_j|C_i)$ was the probability of tone cloud E_j in a given category C_i. Posterior probability $P(C_i|x)$ was given with Bayes rule:

$$P(C_i|x) \propto P(x|C_i)P(C_i). \qquad \text{(Equation 5)}$$

P(C_i) was a prior probability of category i. Each category provided a reward in left or right action a: $r_{a,C}$. The expected reward R of each action for a given sensory stimulus x was as follows:

$$R_{left,x} = P(low|x)r_{left,low} + P(high|x)r_{left,high}$$
$$R_{right,x} = P(low|x)r_{right,low} + P(high|x)r_{right,high}.$$ (Equation 6)

The optimal choice π_x^* was then determined as follows:

$$\pi_x^* = argmax_{a \in \{left,right\}}[R_{a,x}].$$ (Equation 7)

In particular, the choice threshold x_0 satisfied $R_{left,x_0} = R_{right,x_0}$.

In our task, the reward amount of error choice was zero. When the low-category tones were associated with the reward in left choice, the rewards satisfied $r_{left,high} = r_{right,low} = 0$. The prior probabilities of the low and high categories were equal in our task. In these cases, the choice threshold x_0 satisfied the following equation:

$$P(x_0|low)r_{left} = P(x_0|high)r_{right}.$$ (Equation 8)

We wrote $r_{left,low}$ and $r_{right,high}$ as r_{left} and r_{right}. The optimal threshold depended on the perceptual uncertainty and reward size of each choice.

Reinforcement learning model

The reinforcement learning (RL) model assumed that the expected reward of each choice Q_a was updated on a trial-by-trial basis by the choice and outcome:

$$Q_a(t+1) = \begin{cases} Q_a(t) + \alpha(r(t) - Q_a(t)) & if\ a = a(t) \\ (1-\alpha)Q_a(t) & if\ a \neq a(t) \end{cases},$$ (Equation 9)

where α was the learning rate. We used forgetting Q-learning, which fit better in the choice behavior of rodents than a standard Q-learning model in previous studies (Funamizu et al., 2012; Hattori et al., 2019; Ito and Doya, 2015). We then estimated the choice threshold x_0 with a softmax equation and an inverse temperature parameter β:

$$x_0(t) = \frac{\exp(\beta Q_{left}(t))}{\exp(\beta Q_{left}(t)) + \exp(\beta Q_{right}(t))}.$$ (Equation 10)

The softmax equation modeled a perceived reward size that might be different from the actual amount of water (Constantinople et al., 2019; Ito and Doya, 2015). The choice probability in each trial was estimated from a duration-dependent perceptual uncertainty σ_s and a bias parameter d:

$$P(right, t) = \int_{x_0(t)+d}^{1} ZN(x|E_{right}(t), \sigma_{s(t)}^2)dx.$$ (Equation 11)

The initial expected reward for each choice was the average amount of reward (i.e., 2.4). In the RL shared model (Figure 4), one pair of Q_{left} and Q_{right} was updated in both the long- and short-stimulus trials. The RL independent model had separate pairs of Q_a for the long and short trials.

Model comparison

To investigate which behavioral model fit the choices made by the mice, we defined the likelihood $l(t)$ from the estimated choice probability in each trial (Equations 3 and 11):

$$l(t) = \begin{cases} P(right, t) & if\ a(t) = right \\ 1 - P(right, t) & if\ a(t) = left \end{cases}.$$ (Equation 12)

We then analyzed the likelihood in each session L using the trials during the left and right blocks:

$$L = \prod_{t=1}^{T} l(t),$$ (Equation 13)

where T was the number of trials. Model comparison within the psychometric functions used the likelihood ratio test (Daw, 2011), which investigated whether an additional parameter in the model significantly increased the averaged log likelihood per session and determined which parameters were relevant for

choices (Figures 2D, 5D, and S3F). The parameters were fit to achieve the maximum likelihood. Also, we used Bayesian information criterion (BIC) for model comparison (Table S1) (Stoilova et al., 2020):

$$BIC = -2\log(L) + k\log(T), \qquad \text{(Equation 14)}$$

where k was the number of free parameters. The results of model fitting by the likelihood ratio test and BIC were consistent.

For comparison between the psychometric function and RL models, the parameters were estimated with cross validation: we analyzed the likelihood in one session by fitting the parameters with the rest of the sessions (leave-one-session-out cross validation). We then compared the likelihood per trial in all the sessions among models (Figure 4D) (Funamizu et al., 2012; Ito and Doya, 2015).

QUANTIFICATION AND STATISTICAL ANALYSIS

All analyses were performed with MATLAB (MathWorks). For model fitting of psychometric functions, we used the likelihood ratio test (Figures 2D, 5D, and S3F) and Bayesian information criterion (BIC) (Table S1) (Methods details, model comparison). For all the other analyses, we used linear mixed-effects models (MATLAB: fitlme), as we analyzed multiple sessions from each subject (mouse). To compare samples in each session, the fixed and random effects were the samples and subjects, respectively. To compare samples in each trial (Figure 3C), the random effects were the subjects and trials; we tested all the 9 combinations of linear mixed-effects models and determined the best fitting model with BIC. All the statistical details are found in the figure legends, figures, or results.