eNeuro

*Research Article: New Research | Cognition and Behavior*

# Dynamics of visual perceptual decision-making in freely behaving mice

https://doi.org/10.1523/ENEURO.0161-21.2022

**Cite as:** eNeuro 2022; 10.1523/ENEURO.0161-21.2022

Received: 13 April 2021
Revised: 14 February 2022
Accepted: 17 February 2022

# Dynamics of visual perceptual decision-making in freely behaving mice

*2. Abbreviated Title (50 character maximum):* Dynamics of visual decision-making in mice

*3. Author Names and Affiliations:*
**Wen-Kai You[1,2] and Shreesh P. Mysore[1,2]***
[1]Department of Psychological and Brain Sciences, Johns Hopkins University
[2]The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University

*4. Author Contributions:* WKY and SPM designed the research and wrote the paper. WKY performed the experiments, and analyzed the data.

*5. Correspondence should be addressed to:*
Shreesh P. Mysore, PhD
3400 N Charles St, Ames Hall 224
Baltimore MD 21218
mysore@jhu.edu

*6. Number of Figures*: 5

*7. Number of Tables*: None

*8. Number of Multimedia:* None

*9. Number of words for Abstract:* 242

*10. Number of words for Significance Statement:* 77

*11. Number of words for Introduction:* 394

*12. Number of words for Discussion:* 1973

*13. Acknowledgements:* We thank James Garmon for help with fabrication of custom equipment.

*14. Conflict of Interest:* Authors report no conflict of interest.

*15. Funding sources:* This work was supported by funding from NIH R03 HD093995 and startup funds from Johns Hopkins University.

# Dynamics of visual perceptual decision-making in freely behaving mice


## ABSTRACT

Studying the temporal dynamics of perceptual decisions offers key insights into the cognitive processes contributing to it. Conducting such investigation in a genetically tractable animal model can facilitate the subsequent unpacking of the mechanistic basis of different stages in perceptual dynamics. Here, we investigated the time course as well as fundamental psychophysical constants governing visual perceptual decision-making in freely behaving mice. We did so by analyzing response accuracy against reaction time (i.e., conditional accuracy), in a series of 2-AFC orientation discrimination tasks in which we varied target size, luminance, duration, and presence of a foil. Our results quantified two distinct stages in the time course of mouse visual decision-making - a 'sensory encoding' stage, in which conditional accuracy exhibits a classic tradeoff with response speed, and a subsequent 'short term memory (STM)-dependent' stage in which conditional accuracy exhibits a classic asymptotic decay following stimulus offset. We estimated the duration of visual sensory encoding as 200-320 ms across tasks, the lower bound of the duration of short-term memory as ~1700 ms, and the briefest duration of visual stimulus input that is informative as ≤50 ms. Separately, by varying stimulus onset delay, we demonstrated that the conditional accuracy function and RT distribution can be independently modulated, and found that the duration for which mice naturally withhold from responding is a quantitative metric of impulsivity. Taken together, our results establish a quantitative foundation for investigating the neural circuit bases of visual decision dynamics in mice.


## SIGNIFICANCE STATEMENT

This study presents a quantitative breakdown of the time course of visual decision-making in mice during naturalistic behavior. It demonstrates parallel stages in mouse visual perceptual decision dynamics to those in humans, estimates their durations, and shows that mice are able to discriminate well under challenging visual conditions – with stimuli that are brief, low luminance, and small. These results set the stage for investigating the neural bases of visual perceptual decision dynamics and their dysfunction in mice.

## INTRODUCTION

Exploring the temporal dynamics of perceptual decisions from onset of the sensory input through the initiation of behavioral responses affords a key window into the underlying cognitive processes (Uchida, Kepecs et al. 2006, Stanford, Shankar et al. 2010, Siegel, Engel et al. 2011). Investigations of such dynamics in humans (Steinemann, O'Connell et al. 2018, Wilming, Murphy et al. 2020) and other species (Yang, DeWeese et al. 2008, Zariwala, Kepecs et al. 2013, Thura and Cisek 2014) have revealed distinct stages in perceptual processing, their timing, and their interactions. (Wickelgren 1977, McElree and Dosher 1989, Heitz 2014). Performing such investigations in a genetically tractable animal model can additionally facilitate the subsequent unpacking of the mechanistic basis of different stages in perceptual dynamics. However, despite the recent rise in the use of the laboratory mouse for the study of the visual system (Huberman and Niell 2011, Glickfeld, Reid et al. 2014, Seabrook, Burbridge et al. 2017) and of visually guided decision-making (Prusky, West et al. 2000, Prusky and Douglas 2004, Busse, Ayaz et al. 2011, Histed, Carvalho et al. 2012, Carandini and Churchland 2013, Glickfeld, Histed et al. 2013, Long, Jiang et al. 2015, Burgess, Lak et al. 2017, Wang and Krauzlis 2018, Speed, Del Rosario et al. 2020, You and Mysore 2020), the temporal dynamics of visual perceptual decisions represents a significant gap in mouse visual psychophysics (Histed and Maunsell 2014, Umino, Pasquale et al. 2018, Nomura, Ikuta et al. 2019).

In this study, we adapted approaches from human psychophysical studies to investigate the dynamics of visual decision-making in freely behaving mice. In a series of experiments involving touchscreen-based (Mar, Horner et al. 2013, You and Mysore 2020), 2-alternative forced choice (2-AFC) orientation discrimination tasks, we investigated the effect of stimulus size, luminance, duration, delay, and the presence of a competing foil on mouse decision performance (accuracy and reaction time), and importantly,

on the conditional accuracy function. We identified two distinct stages in the time course of mouse visual decision-making within a trial, as has been reported in humans (Posner and Keele 1967, Phillips and Baddeley 1971, Dick 1974, Coltheart 1980, Shibuya and Bundesen 1988, Busey and Loftus 1994, Vogel, Woodman et al. 2006, Bays, Gorgoraptis et al. 2011). In the first 'sensory encoding' stage (Shibuya and Bundesen 1988, Busey and Loftus 1994, Vogel, Woodman et al. 2006, Bays, Gorgoraptis et al. 2011), response accuracy exhibited a classic tradeoff with response speed, and asymptoted to a peak level. In the next stage, response accuracy did not exhibit such a tradeoff, but instead, decayed following stimulus offset, consistent with a classic short-term memory (STM)-dependent process (Posner and Keele 1967, Phillips and Baddeley 1971, Dick 1974, Coltheart 1980). Combining these results with those from drift diffusion modeling (Ratcliff, Smith et al. 2016) allowed us to estimate fundamental psychophysical constants in mouse perceptual decision-making: the time needed by mice to complete visual sensory encoding, the duration for which their short-term memory can intrinsically support discrimination behavior after stimulus input is removed, and the shortest visual stimulus duration that is informative. Additionally, by varying stimulus onset delay, we demonstrated that the two components of accuracy, namely, the conditional accuracy function and the RT distribution can be independently modulated by task parameters. This also allowed a quantitative estimation of impulsivity of mice. Together, this study reveals parallels between mouse and human visual decision dynamics, despite differences in their sensory apparatuses, and enable investigations into the neural circuit underpinnings of the time course of perceptual decision-making in mice.

## METHODS

**Animals.** Thirty-seven mice (33 C57B16/J mice, all male; 4 PV-Cre mice, 3 female, Jackson Labs) were housed in a temperature (~75F) and humidity (~55%) controlled facility on a 12:12h light:dark cycle; ZT0=7 am. All procedures followed the NIH guidelines and were approved by the [Author Institutions] Animal Care and Use Committee (ACUC). Animals were allowed to acclimate for at least one week, with *ad libitum* access to food and water before water regulation was initiated per previously published procedures (Guo, Hires et al. 2014). Briefly, mice were individually housed (for monitoring and control of daily water intake of each identified animal), and administered 1mL water per day to taper their body weight down, over the course of 5-7 days, to 80-85% of each animal's free-feeding baseline weight. During behavioral training/testing, the primary source of water for mice was as a reinforcer for correct performance: 10 µL of water was provided for every correct response. Experiments were all carried out in the light phase.

**Apparatus.** Behavioral training and testing were performed in soundproof operant chambers equipped with a touchscreen (Med Associates Inc.), a custom-built reward port (fluid well), infrared video cameras, a house light and a magazine light above the reward port. The reward port was located at the opposite wall of the chamber relative to the touchscreen (Fig. 1A, 1-1A). Mice were placed within a clear plexiglass tube (5cm diameter) that connects the touchscreen and the reward port. A thin plexiglass mask (3 mm thickness) was placed 3 mm in front of the touchscreen with three apertures (1cm diameter) through which mouse was allowed to interact with the screen via nose-touch. The 'left' and 'right' apertures were placed 3cm apart (center-to-center) along the base of the triangle, and a 'central' aperture, at the apex of the triangle, was 1.5 cm below the midpoint of the base. All experimental procedures were executed using control software (K-limbic, Med-Associates).

**Visual stimuli.** Visual stimuli were bright objects on the dark background (luminance = 1.32 cd/m$^2$). A small cross (60x60 pixels; luminance = 130 cd/m$^2$) was presented in the central aperture and had to be touched to initiate each trial. Oriented gratings (horizontal or vertical) were generated using a square wave, with fixed spatial frequency (24 pixels/cycle) known to be effective for mice to discriminate (Histed, Carvalho et al. 2012). The dark phase of the grating was black, identical to the background (luminance, $L_{dark}$= 1.32 cd/m$^2$), and the bright phase was varied between 1.73 cd/m$^2$ and 130 cd/m$^2$ depending on the tasks (see below). (Note that as the luminance of the bright phase of the grating changed, the contrast of the grating also changed. For clarity, we refer to this stimulus manipulation as a change in luminance,

throughout.) The size of the stimulus was also varied depending on the task, ranging from 60 pixels x 60 pixels to 108 pixels x 108 pixels, which subtended 25-45 visual degrees at a viewing distance of 2 cm from the screen (Fig. 1-1A).

**Experimental procedure and behavioral training.** Each mouse was run for one 30 min behavioral session per day, with each session yielding 80-180 trials. Each trial in a session was initiated by the mouse touching the zeroing cross. Upon trial initiation, the cross vanished, and the visual stimulus (or stimuli) were immediately presented (except in the delay task), for a duration of 0.1-3s depending on the task (see below). Mice were trained to report the orientation of target grating, by nose-touching the correct response aperture (vertical → left; horizontal → right). A correct response triggered a tone (600 Hz, 1 sec), the magazine light turning on, and the delivery of 10μL of water. When mice turned to consumed the reward, their head entry into the reward port was detected by an infrared sensor which caused the zeroing cross (for the next trial) to be presented again. An incorrect response triggered a 5-s timeout, during which the house light and the magazine light were both on and zeroing cross was unavailable for the next trial to be initiated. A failure to respond within 3s (starting stimulus presentation) resulted in a trial reset: the stimulus vanished and the zeroing cross was presented immediately (without a timeout penalty), to allow initiation of the next trial. Well-trained animals failed to respond on fewer than 5% of the total number of trials, and there were no systematic differences in the proportion of such missed trials between different conditions. Within each daily 30-minute behavioral session, mice consumed approximately 1mL of water. If a mouse failed to collect enough water from the behavioral session, they were provided with a water supplement using a small plastic dish in their home cage.

**Single-stimulus discrimination task.** Upon trial initiation, a single grating stimulus (i.e., the 'target') was presented above the central aperture, at the same horizontal level as the left and right apertures, and mice were required to report its orientation with the appropriate nose-touch (Fig. 1B). When stimulus size and luminance were manipulated (Fig. 1, and 2), three different sizes were tested: 60x60, 84x84, 108x108 (pixels x pixels). For each size, seven different levels of luminance were tested: 2.00, 2.59, 4.37, 7.55, 16.2, 34.3, 130 cd/m$^2$. (These corresponded nominally to Michelson's contrasts of 20%, 32%, 54%, 70%, 85%, 93%, 98%, respectively; Michelson's contrast is computed as ($luminance_{bright}$ - $luminance_{dark}$) / ($luminance_{bright}$ + $luminance_{dark}$) *100.) Trials with different stimulus luminance at a particular size were interleaved randomly throughout a session, while trials with different stimulus sizes were examined on different days. When the stimulus duration was manipulated (Fig. 3), the luminance (130 cd/m$^2$) and size (60 pix x 60 pix) of the grating were fixed, and eleven different stimulus durations were tested: 100 ms, 200, 300, 400, 500, 600, 800, 1000, 1500, 2000, 3000 ms. The stimulus duration was fixed for a given day, and across days, was varied in a descending sequence from 3000 ms to 100 ms. When the stimulus onset delay was manipulated (Fig. 5), the luminance (130 cd/m$^2$), size (60 pix x 60 pix), and duration (600 ms) of the grating were fixed. Three different delays were tested: 0, 100, and 200 ms. The delay duration was fixed for a given day, and varied in an ascending sequence from 0 ms to 200 ms.

**Flanker task.** Upon trial initiation, either one stimulus ('target', 60 pix x 60 pix, luminance = 20.1 cd/m$^2$) was presented at the lower location, or two stimuli were presented simultaneously, with the target at the lower location and a second 'flanker' at the upper location (Fig.4A). Flankers were of the same size (60 pix x 60 pix) and spatial frequency (24 pixel/cycle) as the target, but with luminance ranging (over 8 levels) from less than that of the target to greater than that of the target (You and Mysore 2020). The orientation of the flanker was either identical to that of the target ('congruent trial') or orthogonal to that of the target ('incongruent trial'). The stimulus (stimuli) was (were) presented for a duration of 1s, and mice were required to report orientation of the target grating with the appropriate nose-touch (within 3s). All types of trials (no flanker, congruent, incongruent) and flanker luminance were interleaved randomly within each daily session. Data from this experiment have been reported previously (You and Mysore 2020), and were re-analyzed here using different analyses.

**Subject inclusion/exclusion.** A total of 37 mice were used in this study, with different subsets used in different tasks. For mice involved in more than one task, they were well-rested for 3-8 weeks with food and water *ad libitum* between experiments. Before the start of each experiment, all mice were given a few days of practice session to ensure that they remembered/re-learned the association between the orientation of single target and the appropriate nose-touch. Of the total of 37 mice trained across tasks, 28 mice passed the inclusion threshold of response accuracy >70% in the single stimulus discrimination task, and were included for the analyses reported in this paper.

**Trial inclusion/exclusion.** Mice were observed to become less engaged in the task towards the end of a behavioral session, when they had received a sizeable proportion of their daily water intake. This was reflected in their behavioral metrics: they tended to wait longer to initiate the next trial, and their performance deteriorated. We identified and excluded such trials following a published procedure (You and Mysore 2020), in order to minimize confounds arising from loss of motivation towards the end of sessions. Briefly, we pooled data across all mice and all sessions, treating them as coming from one session of a single 'mouse'. We then binned the data by trial number within the session, computed the discrimination accuracy in each bin (% correct), and plotted it as a function of trial number within session (Fig. 1-1B, 3-1A, 5-1A). Using a bootstrapping approach, we computed the 95% confidence interval for this value. We used the following exclusion criterion: Trials q and above were dropped if the $q^{th}$ trial was the first trial at which *at least one* of the following two conditions was satisfied: (a) the performance was statistically indistinguishable from chance on the $q^{th}$ trial and for the majority (3/5) of the next 5 trials (including the $q^{th}$), (b) the number of observations in $q^{th}$ trial was below 25% of the maximum possible number of observations for each trial (Σ mice*sessions), thereby signaling substantially reduced statistical power available to reliably compare performance to chance. The plots of performance as a function of trial number, and number of observations as a function of trial number for the different tasks in this study are shown in Figs. 1-1B, 3-1A, 5-1A, along with the identified cut-off trial numbers (q).

**Behavioral measurements:** Response accuracy (% correct) was calculated as the number of correct trials divided by the total number of trials responded (correct plus incorrect). Reaction time (RT) was defined as the time between the start of stimulus presentation and time of response nose-touch, both detected by the touchscreen. In the experiment involving stimulus onset delays (Fig. 5A), RT was computed with respect to trial initiation (as opposed to from stimulus onset).

**Drift diffusion modeling of RT distributions.** The RT measured here represents the duration from stimulus onset to completion of execution of the motor response. In order to specifically isolate the time spent in decision making (separately from the latency of activation of sensory neurons as well as duration of motor execution), we applied the drift-diffusion model to our RT data (Voss, Nagler et al. 2013, Voss, Voss et al. 2015). This model hypothesizes that a subject ('decision maker') collects information from the sensory stimulus via sequential sampling, causing sensory evidence to accrue for or against a particular option (usually binary) while viewing the stimulus. A decision is to be made when the accumulating evidence reaches an internal threshold of the subject. This process of evidence accumulation, together with the processes of sensory encoding and motor execution, as well as threshold crossing, determine the RT observed on each trial.

We used a standard version of the model that consists of four independent variables (Ratcliff 1978, Ratcliff, Smith et al. 2016): (1) the drift rate, (2) the boundary separation, (3) the starting point, and a (4) non-decisional constant ($t_{delay}$), which accounts for the time spent in sensory encoding and motor execution. In the case of our tasks, there was no reason for the drift rate to be different between vertical versus horizontal gratings, and therefore, we merged both type of trials (trials with a horizontal target grating and trials with a vertical target grating). We treated 'correct' response and 'incorrect' response as the two binary options, and fit the diffusion model to the RT distributions of correct versus incorrect trials using the fast-dm-30 toolbox with the maximum likelihood option to gain estimates of those four parameters for each individual mouse (Fig. 2-2)(Voss, Voss et al. 2015).

**Conditional accuracy analysis.** Conditional accuracy was calculated as the percentage of correct trials (accuracy) as a function of RT. For this analysis, trials from all mice were pooled together and treated as if they were from one single mouse for statistical power (Fig. 2 onwards; for completeness, conditional accuracy plots using non-pooled data, i.e., from individual mice, are included in Extended Figures). Pooled trials were then sorted by their RT, and then binned by RT such that there were: (1) sufficient number of trials in each bin; and (2) sufficient number of total bins, to ensure the robustness of curve fitting and therefore the estimates of quantitative metrics (see below). Typical bin sizes used were 50ms, 100 ms or 200 ms bins, depending on the experiments and stage of analysis (sensory encoding or STM-dependent). The effect of bin size on the estimates of quantitative metrics is explored in the Extended Figures; results show that the estimates are comparable across tested bin sizes.

**Conditional accuracy function (CAF).** To quantitatively describe the relationship between the conditional accuracy and RT, we fitted the plot of accuracy against binned RT with parametric functions (the CAF; see below) using a nonlinear least square method For RT bins aligned to stimulus onset (Fig. 2, 4C, 5B), we fit the conditional accuracy data using an increasing asymptotic function:

$$\text{Conditional accuracy} = \lambda\,(1 - e^{-\gamma_{enc}\,(RT-\delta)})).$$

Three key metrics were defined for this sensory encoding phase, for use in subsequent comparisons across conditions: (1) peak conditional accuracy ($a_{peak}$), the maximal level of accuracy that the CAF reaches within the range of RT; (2) the slope parameter ($\gamma_{enc}$); and (3) the first instant at which the conditional accuracy reaches its maximal value ($t_{peak}$) - defined as the time point at which the ascending CAF reaches 95% of $a_{peak}$ . Note that $t_{peak}$ is influenced by the peak conditional accuracy ($a_{peak}$), the slope parameter, $\gamma_{enc}$, and the temporal offset at chance performance, $\delta$ For RT bins aligned to stimulus offset (Fig. 3C, 4E, 5D), we fit the decaying conditional accuracy data using a sigmoidal function:

$$\text{Conditional accuracy} = \lambda\,[1/(1 + e^{-\beta_{dec}\,(RT-\,\tau)})]+50$$

Three key metrics were defined for this STM-dependent stage for use in subsequent comparisons across conditions: (1) peak conditional accuracy ($a_{peak}$), the maximal level of accuracy within the range of RT; (2) the first instant ($t_{decay}$) at which conditional accuracy is lower than the maximum - defined as the time point at which the descending CAF crosses 95% of $a_{peak}$; and (3) the first instant ($t_{chance}$) at which conditional accuracy drops to chance levels - defined as the timepoint at which the descending CAF crosses 52.5%. In (rare) cases when the CAF never went below 52.5%, $t_{chance}$ was set to be the upper bound of the window of analysis (i.e., 3000ms – stimulus duration = the window for which the mice can respond following stimulus offset). Note that $t_{decay}$ and $t_{chance}$ are influenced by both the slope parameter, $\beta_{dec}$, and $\tau$.

Confidence intervals of the CAF fits, as well as for the parameters, were estimated by standard bootstrapping procedures involving resampling the raw data randomly with replacement (1000 x), to get repeated estimates of the CAF and corresponding metrics. In all relevant figures, the box plots of the estimated values of each metric show the median (the central mark), the 25th and 75th percentiles (the bottom and top edge of the box), and the most extreme data points not considered as outliers (whiskers).

In the experiment in which the stimulus onset delay was manipulated (Fig. 5), we adopted the following two adjustments to our procedure for the analysis of the conditional accuracy function. First, since the stimulus was short (600 ms), in order to ensure robust estimates of CAF metrics for the sensory encoding stage, we included data beyond stimulus offset as well for the fitting of the CAF through 400 ms following offset. (We chose to include data upto 400 ms after offset, specifically, because we had learned from Figure 3 that conditional accuracy remains at its plateau for nearly 500 ms following stimulus offset.) Second, we also excluded trials with RT < 200ms for the fitting of the CAF (Fig. 5B), because these represent trials on which responses were initiated prematurely (200 ms represents our estimate of the duration of sensory latency plus motor execution; see text surrounding Figure 2).

**Statistical tests**. All analyses and statistical tests were performed in MATLAB. For single-stimulus experiments in which only one stimulus parameter was systemically varied, one-way ANOVA was applied to examine the effect of the manipulating the single factor (duration and delay, Fig. 3AB, 5A, 1-1CD). For experiments that involved changing both stimulus size and luminance (Fig. 1CDE, 2-2), two-way ANOVA was applied to examine the effect of each factor, as well as their interaction. For the flanker task, the paired-sample t-test was used to examine if the group performance was different between trial types (Fig. 4B).

For the metrics associated with CAF, comparisons were made by measuring the effect size (Hedges' g) of the difference between two distributions (Fig. 2BD, 4DF and 5CE). All effect size measurements, including those with ANOVA ($\eta^2$), were calculated following the methods (and source code) of Hentschke and Stüttgen (2011)(Hentschke and Stüttgen 2011). Hedges' g estimates the distance between the two distributions in units of their pooled standard deviation, with larger numbers indicating stronger effects. $\eta^2$ varies from 0 to 1, with larger values indicating greater ratio of variance explained in the dependent variable by a predictor while controlling the other variables.

## RESULTS

In this study, freely behaving mice were trained to perform 2-AFC orientation discrimination in a touchscreen-based setup (Mar, Horner et al. 2013, You and Mysore 2020)(Methods). Briefly, mice were placed in a plexiglass tube within a soundproof operant chamber equipped with a touch-sensitive screen at one wall and a reward well at the opposite wall (Fig. 1A). A plexiglass sheet with three holes was placed in front of the touchscreen - the holes corresponded to the locations at which mice were allowed to interact with the screen by a nose-touch (Fig. 1A). All trials began with a nose-touch on a bright zeroing-cross presented within the lower central hole (Fig. 1B). Immediately following nose-touch, an oriented grating (target; bright stimulus on a dark background) was presented at the center of the screen. Mice were rewarded if they responded to the orientation of the target with an appropriate nose-touch: vertical (horizontal) grating → touch within upper left (upper right) hole. Behavioral data were collected from daily sessions that lasted 30 minutes for each mouse.

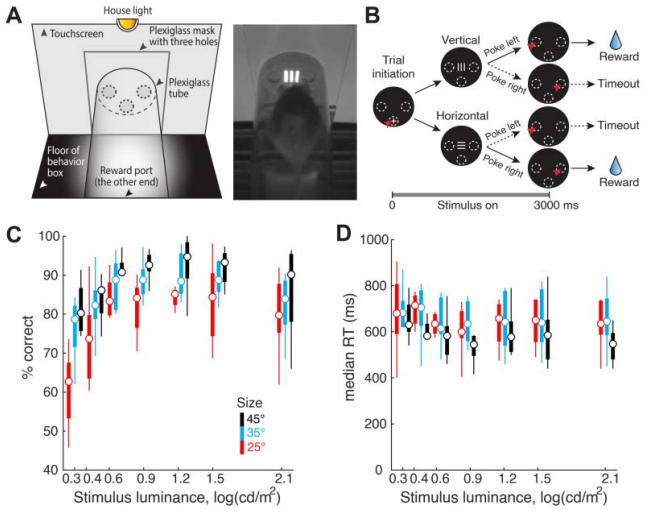


**Figure 1. Stimulus luminance and size modulate orientation discrimination performance in freely behaving mice. (A)** Left: Schematic of touchscreen-based experimental setup showing key components. Right: Snapshot of freely behaving mouse facing a visual stimulus on the touchscreen. **(B)** Schematic of 2-AFC task design. Black discs: Screenshots of touchscreen with visual stimuli; dashed ovals: locations of holes through which mice can interact with

touchscreen; white '+': zeroing cross presented within central response hole at start of each trial; red arrowhead: nose-touch by mouse. Shown also are vertical or horizontal grating stimuli, and reinforcement (water)/punishment (timeout) schedule. Bottom: Trial timeline. 0 ms corresponds to the instant at which the mouse touches the zeroing cross (trial initiation). Immediately following this, the target grating was presented and stayed on for 3s, or until the mouse responded, whichever came first. Vertical and horizontal targets were interleaved randomly. **(C)** Psychometric plots of discrimination accuracy against stimulus luminance (n=8 mice). Different colors correspond to different target sizes. 2-way ANOVA, $p<0.001$ (luminance), $p<0.001$ (size), $p=0.498$ (interaction). Effect size $\eta^2=0.292$ (luminance), $\eta^2=0.192$ (size), $\eta^2=0.037$ (interaction). For each stimulus size/luminance, the box plot shows the median (the central mark), and the 25$^{th}$ and 75th percentiles (the bottom and top edge of the box) of the group (n=8). The whiskers extend to the most extreme data points not considered as outliers. **(D)** Plot of median reaction time (RT) against stimulus luminance. 2-way ANOVA, $p=0.998$ (luminance), $p=0.004$ (size), $p=1$ (interaction). Effect size $\eta^2=0.003$ (luminance), $\eta^2=0.071$ (size), $\eta^2=0.010$ (interaction).
**See also Fig. 1-1.**

**Stimulus size and luminance modulate mouse discrimination performance**

We first examined the effect of target size and luminance on the decision performance of mice in the orientation discrimination task. Here, the target grating was presented for up to 3 seconds after trial initiation (Fig.1B; Methods), and its size and luminance were systematically varied; the spatial frequency was fixed at 0.1 cycles/degree (24 pixels/cycle) (Prusky and Douglas 2004, Histed, Carvalho et al. 2012) (Methods). Mice were allowed to respond at any time during stimulus presentation, and the stimulus was terminated automatically upon response.

We found that both the target luminance and size significantly modulated discrimination accuracy (Fig. 1C, 2-way ANOVA, main effect of luminance, $p<0.001$, effect size $\eta^2=0.292$; main effect of size, $p<0.001$, $\eta^2=0.192$; interaction, $p=0.498$, $\eta^2=0.037$). These results revealed that mice discriminated target orientation better than chance even at the lowest luminance (2.00 cd/m$^2$) and size (25°) tested (Fig. 1C; the red box at the left lower corner, $p=0.015$, *t*-test against mean accuracy=50%, effect size g1=1.129). Additionally, at this smallest target size (25º), mice could discriminate with >80% accuracy for most of the tested luminance values (≥4.37 cd/m$^2$; Fig. 1CD, red data).

The effect of these parameters on median reaction times (RTs) was less pronounced. Target size, but not luminance, modulated reaction times (RTs) (Fig.1E, two-way ANOVA; main effect of size, $p=0.004$, effect size $\eta^2=0.071$; main effect of luminance, $p=0.998$, $\eta^2=0.003$; interaction, $p=1$, $\eta^2=0.010$). Together, these results revealed a systematic effect of target size and luminance on discrimination accuracy.

**Effect of stimulus size and luminance on dynamics of visual decision-making: the sensory encoding stage**

To investigate the dynamics of visual perceptual decision-making, we adapted approaches from human studies and examined the dependence of response accuracy on RT, i.e., the so-called 'conditional accuracy' function (CAF) (Wickelgren 1977, McElree and Dosher 1989, Heitz 2014). For these analyses, we pooled trials from all mice (n=8) in order to gain better statistical power for the estimates of parameters of the CAF (Methods; plots of the data for individual mice showed similar overall shapes of the CAF; Fig.2-1A).

Specifically, we investigated the dynamics of visual perceptual decision-making as a function of stimulus size, and separately, as a function of stimulus luminance. First, to examine the effect of stimulus size on decision dynamics, we pooled trials from all mice across luminance values (7 luminance values) for each stimulus size, sorted them based on RT, and plotted conditional accuracy for each RT bin (100ms; Fig. 2A; Methods). We found that for responses with RT less than ~500 ms, conditional accuracy improved for longer RT, consistent with the classic 'speed-accuracy tradeoff' [34]. For responses with RT greater than 500 ms and up to 3s, the allowed duration for responses, conditional accuracy plateaued, and was independent of RT. Next, to examine the effect of stimulus luminance on decision dynamics, we pooled

trials from all mice across size values into two groups based on stimulus luminance: (1) trials with target luminance ≤ 4.37 cd/m$^2$ ('low luminance'), and (2) trials with target luminance > 4.37 cd/m$^2$ ('high luminance'; Methods). Here, as well, we found a similar initial stage of increasing conditional accuracy upto RT of ~ 500 ms, followed by a plateauing of conditional accuracy.

Drawing upon arguments from human behavioral studies, we reasoned that the initial transient stage of the conditional accuracy function reflects the process of sensory encoding: during it, slower responses allow more sensory evidence to be acquired, thereby improving conditional accuracy up to a peak value reflecting the completion of sensory encoding (Shibuya and Bundesen 1988, Busey and Loftus 1994, Vogel, Woodman et al. 2006, Bays, Gorgoraptis et al. 2011).

To quantify these dynamics, we fit the conditional accuracy data with an asymptotic function (Fig. 2AC, solid curves) (Wickelgren 1977, McElree and Dosher 1989, Heitz 2014), and estimated three key metrics, in each case: (1) the peak conditional accuracy ($a_{peak}$), (2) the slope parameter ($\gamma enc$), and (3) the timepoint at which conditional accuracy reached its peak ($t_{peak}$; Methods).

We found that the peak conditional accuracy was significantly modulated by stimulus size (Fig.2B-left; $a_{peak}$: size 25°, median [C.I.] = 81.3 [79.1, 83.7] %; size 35º = 88.0 [86.5, 89.4] %; size 45º = 92.4 [90.7, 94.1] %; effect size Hedge's g= -6.71 (25°-35°), -5.39 (35°-45°), -10.6 (25°-45°)), but not the slope of the function (slope parameter, $\gamma enc$, Fig. 2B-middle, size 25° = 6.52 [5.10, 9.07] a.u.; size 35° = 8.81 [7.09, 10.6] a.u.; size 45° = 7.92 [6.15, 10.1] a.u. Hedges' g= -2.24 (25°-35°), 0.863 (35°-45°), -1.34 (25°-45°)), or the time to reach peak accuracy ($t_{peak}$, Fig. 2B-right, size 25° = 493 [375, 597] ms; size 35° = 459 [420, 505] ms, size 45° = 466 [420, 522] ms; Hedges' g= 0.728 (25°-35°), -0.274 (35°-45°), 0.558 (25°-45°)).

Next, we found that the peak conditional accuracy was higher in high-luminance trials (Fig. 2D-left, low-luminance, median [C.I.] = 84.7 [82.7, 86.3] %; high-luminance = 89.5 [88.2, 90.7] %, effect size Hedges' g= -6.13). The slope was also higher in high-luminance trials (slope parameter, $\gamma enc$, Fig. 2D-middle, low-luminance = 6.37 [5.21, 7.78] a.u.; high-luminance = 10.32 [8.49, 12.6] a.u., Hedges' g= -4.51) suggesting a faster rate of sensory encoding in high-luminance trials. Consistent with this, the time to reach peak accuracy was shorter in high-luminance trials (Fig.2D-right; $t_{peak}$: low-luminance = 531 [478, 599] ms; high-luminance = 412 [378, 448] ms, Hedges' g= 4.86).

The RT measured here represents the duration from the start of the sensory input to the completion of motor response. In order to obtain an estimate of the duration, specifically, of decision-making, we employed the standard drift diffusion modeling (DDM) approach (Ratcliff 1978, Ratcliff, Smith et al. 2016) (Methods). Briefly, the DDM analyzes the full RT distribution and yields a quantitative estimate of four parameters (Methods), one of which is $t_{delay}$, a parameter which accounts for the combination of: (a) the time taken for the sensory (visual) periphery to transduce and relay information to visual brain areas (i.e., neural response latency), as well as (b) the time taken for executing the motor response (i.e., motor execution delay). In our tasks, the latter corresponds to the time for the mouse to move its head (and body) to achieve the appropriate nose-touch.

Using this approach, we found that stimulus size as well as luminance had no discernable effect on $t_{delay}$ (Fig. 2-2. 2-way ANOVA, size: p=0.308, luminance: p=0.523; interaction: p=0.931), and the average value of $t_{delay}$ was 212 ms. Consequently, we estimated the duration of just the sensory encoding stage (temporal integration window) as $t_{peak} - t_{delay} = t_{peak} - 212$ ms. Across conditions, this took values of 200 ms (412 ms -212 ms; high luminance), 254 ms (466-212 ms; size of 45 deg), 247 ms (459-212 ms; size of 35 deg), 281 ms (493-212 ms; size of 25 deg), and 319 ms (531 ms -212 ms, low luminance).

Thus, conditional accuracy analysis allowed us to quantify the sensory encoding stage in mouse visual perceptual dynamics. We estimated its duration to be brief, varying between 200 ms and 320 ms across the tested conditions.

Following the completion of sensory encoding, a fully constructed representation of the sensory stimulus is available, as a result of which, additional sampling of the stimulus brings no extra benefits to the performance. Our finding that RTs longer than $t_{peak}$ produce no further increase in conditional accuracy, is consistent with the view (Fig. 2AC).

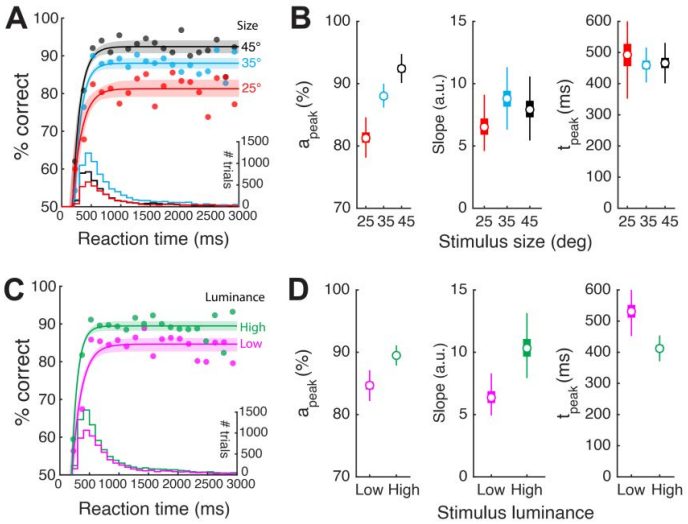


**Figure 2. Stimulus size and luminance modulate the sensory encoding stage of the conditional accuracy function (CAF). (A)** Plot of accuracy as a function of RT bins (conditional accuracy) using same dataset as Fig. 1. Data pooled across all stimulus luminance and mice (n=8), sorted by stimulus size; RT bin size = 100 ms. Solid curves: Conditional accuracy functions (CAFs, best-fit rising asymptotic function; Methods) for targets of different sizes (black: 45º; blue: 35º; red: 25º); light shading: 95% CI of the fit (Methods). Histograms at bottom: RT distributions for targets of different sizes (y-axis on the right). The overall response accuracy for a particular stimulus condition is the dot product of the CAF and the RT distribution. **(B)** Box plots of the key parameters for different target sizes. Left panel: $a_{peak}$; middle panel: slope parameter; right panel: $t_{peak}$. **(C)** CAFs for targets of different luminance conditions (magenta: 'low' luminance - first three luminance levels from Fig. 1C; green: 'high' luminance - last four luminance levels; Methods); conventions as in A. **(D)** Box plots of the key parameters for different luminance conditions; conventions as in C. The box plots in all panels show the median (open circle), the 25th and 75th percentiles (the bottom and top edge of the box), and the most extreme data points not considered as outliers (whiskers); in some panels, the boxes are the same size as the symbol for the median.
**See also Fig. 2-1, 2-2.**

**Stimulus duration and the dynamics of visual decision-making: the memory-dependent stage**

The next stage in the time course of perceptual decisions has been identified in human studies as the so-called 'short-term memory' (STM)-dependent stage, during which an internal representation of the sensory stimulus is available transiently in memory for guiding behavior (Smith and Ratcliff 2009). Studies have demonstrated the STM to be labile such that once the stimulus is terminated, sensory information maintained in STM decays and is lost (over seconds) (Brown 1958, Gold, Murray et al. 2005, Zhang and Luck 2009, Barrouillet and Camos 2012, Ricker, Vergauwe et al. 2016).

In our experiments so far, the target stimulus was present on the screen for the full duration of the response window (3s). Here, in order to investigate and quantify the STM-dependent stage of mouse perceptual decisions, we performed an experiment in which we shortened the stimulus duration systematically from 3s to 100 ms. This allowed us to examine the time course of decision behavior following stimulus offset, and, as well, to examine the shortest stimulus that mice are able to discriminate effectively.

We first examined overall mouse behavioral performance at different stimulus durations. We found that response accuracy was significantly modulated (Fig.3A, one-way ANOVA, $p<0.001$, effect size $\eta^2=0.331$), with accuracy decreasing for shorter stimulus durations (Pearson's $\rho=0.712$, $p=0.014$). There was also a trend of decreasing median RT for shorter stimulus durations (Fig.3B, one-way ANOVA, $p=0.056$, effect size $\eta^2=0.177$; Pearson's $\rho=0.861$, $p=0.001$). Additionally, these results revealed, that the shortest stimulus duration needed for mice to be able to discriminate above chance was less than 100 ms - the smallest duration tested (Fig. 3B).

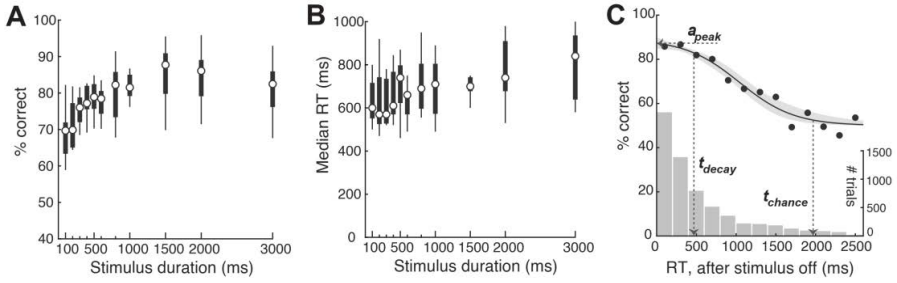


**Figure 3. Stimulus duration and the memory-dependent stage of the conditional accuracy function.**
**(A)** Psychometric plot of discrimination accuracy against stimulus duration (n=9 mice; 1-way ANOVA; $p<0.001$. effect size $\eta^2=0.331$). **(B)** Plot of median reaction time (RT) against stimulus duration (1-way ANOVA; $p=0.056$. effect size $\eta^2=0.177$). **(C)** Plot of the conditional accuracy (solid data) as a function of RT bins relative to stimulus offset. Only trials in which the stimulus was longer than 332 ms were included (in order to ensure full sensory encoding - see text; Methods). Curve and shading: best-fit sigmoid function and 95% C.I. Bootstrapped estimates of each key metric: $a_{peak}$, median [C.I.] =87.3 [84.8, 89.9] %; $t_{decay}$ = 469 [279, 697] ms; and $t_{chance}$ = 1969 [1708, 2520] ms. Histogram: RT distribution (y axis on the right). In this experiment, stimulus size and luminance were maintained fixed at 25°and 130 cd/m$^2$respectively.
**See also Fig. 3-1.**

Next, to examine the decision dynamics following stimulus offset, we aligned trials to stimulus offset, and computed the conditional accuracy. Considering that incomplete sensory encoding may be a confounding factor to the STM decay, we only included those trials on which the stimulus was presented for longer than the duration of the sensory encoding stage, estimated in Figure 2 to be 320 ms.

We observed the classic decay in conditional accuracy with longer RTs (Fig. 3C). To quantify the time course of the decay, we fit the conditional accuracy data with a sigmoidal function (Methods), and estimated three key metrics (Fig. 3C; Methods). The first, peak performance, $a_{peak}$, was 87.3% (median, C.I.= [84.8, 89.9] %), comparable to the asymptotic level of Figure 2, thereby supporting that sensory encoding is, indeed, complete on these trials. The second, the time point at which the conditional accuracy dropped below the peak value, $t_{decay}$, was 469 ms (median, C.I.= [279, 697] ms) after stimulus offset. The third, the first timepoint at which the discrimination accuracy dropped to a level indistinguishable from the chance, $t_{chance,}$ was 1969 ms (median, C.I.= [1708, 2520] ms) after stimulus offset (Methods).

Thus, our conditional accuracy analysis allowed us to investigate quantitatively the second, STM-dependent stage in mouse visual perceptual dynamics. We estimated the duration over which above-chance decision accuracy is supported in mice after stimulus offset as ~1700ms (i.e., $t_{chance}$ minus the $t_{delay}$).

**The presence of flanker stimulus modulates perceptual dynamics**

We next investigated the impact of sensory context on visual decision dynamics. It is well-established that the sensory context in which the perceptual target is presented modulates animals' behavior (Miller 1991, Meier, Flister et al. 2011, Whitney and Levi 2011). For instance, in the classic flanker task in humans, the co-occurrence of a flanker stimulus with conflicting information can interfere with perceptual performance (Eriksen and Eriksen 1974, Fan, McCandliss et al. 2002). Recently, similar results were demonstrated in mice using a touchscreen version of the flanker task (You and Mysore 2020). In this task (Fig. 4A), a target grating (always presented at the lower location) was accompanied by a flanker grating at the upper location with either orthogonal orientation ('incongruent' flanker) or same orientation ('congruent' flanker). Compared to the presence of a congruent flanker, the 'incongruent' flanker significantly impaired discrimination accuracy (Fig. 4B-left; $p<0.001$, paired-sample *t* test. effect size Hedges' g=1.61; re-plotted based on data from (You and Mysore 2020); Methods). Here, we analyzed that dataset with the conditional accuracy analysis to investigate whether an incongruent flanker affected the sensory encoding stage or the STM-dependent stage of perceptual dynamics.

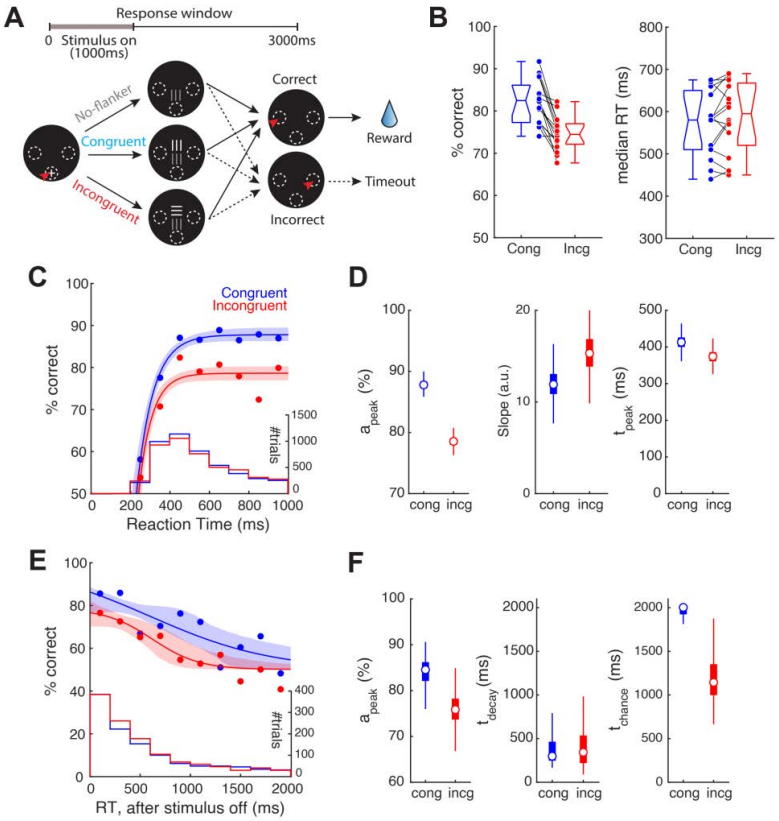

**Figure 4. Incongruent flanker modulates the sensory encoding stage of the conditional accuracy function (CAF).** **(A)** Schematic of the flanker task; target grating is always presented at the lower location; a second 'flanker' grating (orthogonal orientation – incongruent flanker, or same orientation – congruent flanker) is presented simultaneously, and always at the upper location; luminance of flanker is systematically varied (adapted from (You and Mysore 2020)). All other conventions as in Figure 1. The stimuli were presented for 1s and the response window was 3s. **(B)** Left panel: Comparison of performance between trials with incongruent vs. congruent flanker. $p<0.001$, paired-sample *t* test. effect size Hedges' g=1.61. Right panel: Comparison of median RT between trials with incongruent vs. congruent flanker. p=0.137, paired-sample *t* test. effect size Hedges' g=-0.176. Data re-analyzed from You et al (You and Mysore 2020); each line represents data from one mouse (n=17 mice). Data in B-F include only trials with high flanker luminance (≥20.1 cd/m$^2$; see text). **(C)** CAFs of the sensory encoding stage; Blue: trials with congruent flanker; red: trials with incongruent flanker; histograms; RT distributions. **(D)** Key parameters of CAFs (sensory encoding stage) for trials with congruent vs. incongruent flanker; $a_{peak}$ (left), slope parameter (middle), and $t_{peak}$ (right). Box plots show the distribution of bootstrapped estimates (Methods). Effect sizes (congruent – incongruent): $a_{peak}$: Hedges' g=11.0; slope parameter: Hedges' g=-1.73; $t_{peak}$: Hedges' g=2.08. Note, the sizes of the boxes in the left and right panels are similar to the sizes of the circular symbols depicting the medians. **(E)** CAFs of the STM-dependent stage; data aligned to stimulus offset. Blue: trials with congruent flanker; red: trials with incongruent flanker. **(F)** Plots of key parameters of CAFs (STM-dependent stage) for trials with congruent vs. incongruent flanker; $a_{peak}$ (left), $t_{chance}$ (middle) and $t_{decay}$ (right). Conventions and statistical methods as in D. $a_{peak}$: Hedges' g=2.54; $t_{chance}$: Hedges' g=2.98; $t_{decay}$: Hedges' g=0.175.

To investigate the effect of the flanker on perceptual dynamics, we pooled trials from all mice into two groups based on their flanker congruency, and sorted the trials based on their RT. Since previous study (You and Mysore 2020) has demonstrated that the flanker affects performance significantly only when its luminance is higher than (or equal to) that of the target, here we included only high-luminance trials (trials with flanker luminance ≥20.1 cd/m$^2$). To examine the sensory encoding stage quantitatively, we followed the approach used in Figure 2 and selected the trials on which mice responded before the stimulus ended (RT < 1000ms), and aligned them to stimulus onset. Separately, to examine the STM-dependent stage, we followed the approach used in Figure 3 and selected the trials on which responses were made after the stimulus ended, and aligned them to stimulus offset.

The sensory encoding stage was significantly modulated by flanker congruency (Fig. 4CD). We found that, the peak conditional accuracy for incongruent trials was significantly lower than that of congruent trials (Fig. 4D-left; $a_{peak}$: congruent, median [C.I.] = 87.8 [86.3, 89.6] %, incongruent = 78.5 [76.9, 80.2] %; effect size (congruent-incongruent) Hedges' g=11.0), indicating that the presence of a high-luminance incongruent flanker interfered with the sensory encoding of the target stimulus. While the slope parameter for incongruent trials remains comparable to that of the congruent trials (Fig. 4D middle; congruent = 11.9 [9.10, 15.5] a.u., incongruent = 15.3 [11.6, 20.6] a.u.; Hedges' g=-1.73), the time to reach peak accuracy was, however, shorter for incongruent trials (Fig. 4D-right; $t_{peak}$: congruent = 413 [378, 458] ms, incongruent = 374 [340, 410] ms; Hedges' g=2.08), consistent with the lower $a_{peak}$ (Fig. 4D-left).

The STM-dependent stage also appeared to be modulated by flanker congruency (Fig. 4EF). Following stimulus offset, the time at which conditional accuracy dropped to chance was much earlier in incongruent trials than in congruent trials (Fig. 4F-right; $t_{chance}$: congruent, median [C.I.] = 2000 [1363, 2000] ms; incongruent = 1145 [816, 1985] ms; Hedges' g=2.98). However, this was likely due largely to the lower peak conditional accuracy for incongruent trials (Fig.4F-left; $a_{peak}$: congruent= 84.5 [76.9, 88.6] %; incongruent = 75.9 [70.1, 82.1] %; Hedges' g=2.54), as opposed to changes in $t_{decay}$ (Fig. 4F-middle; congruent= 299 [197, 1086] ms; incongruent= 343 [126, 802] ms; Hedges' g=0.175), or to the rate of decay (slope parameter; data not shown, congruent= -1.82 [-10, -1.0] a.u., incongruent= -4.82 [-10.0, -1.60] a.u.; Hedges' g=0.99).

In sum, we found that the presence of an incongruent flanker interferes the sensory encoding stage but not the STM-dependent stage of mouse visual decision dynamics.

**Stimulus onset delay modulates RT distribution but not the conditional accuracy function**

The components of behavioral performance that we have investigated thus far, namely, overall decision accuracy, RT distribution and conditional accuracy function are related formally in the following way: the overall decision accuracy is the dot product of the conditional accuracy function and RT distribution.

Our manipulations, thus far, produced changes in the conditional accuracy function predominantly. Here, we wondered whether task parameters could, instead, alter RT distribution, and possibly do so without affecting the conditional accuracy function. To test this, we added a delay between trial initiation and target onset (called stimulus onset delay) in the single stimulus discrimination task. We reasoned that the extent to which mice are unable to adaptively withhold responding could impact the RT distribution.

We found that adding a stimulus onset delay did alter the RT distribution of mice (Fig. 5A-upper panel). The median RTs, measured relative to trial initiation, showed an increasing trend with delay (one-way ANOVA, p=0.094; effect size $\eta^2$=0.179; Pearson's correlation=0.422, p=0.028). This indicated that mice were able to sense the delayed onset of stimulus and thereby withhold their responses. However, mice were unable to withhold responding for the full duration required. By performing a linear regression (Fig. 5A-upper panel; dashed line), we found that mice were able to withhold their responses for only 39 ms for every 100ms of delay. Separately, this increase in RT for longer delays was accompanied by a trend towards lower decision accuracy (Fig. 5A-lower panel, one-way ANOVA, p=0.182; effect size $\eta^2$=0.132; Pearson's correlation=-0.358, p=0.067).

By contrast, conditional accuracy analysis revealed no effect of stimulus onset delay either on the sensory encoding stage (Fig. 5BC, $a_{peak}$: no-delay, median [C.I.] = 84.2 [82.2, 86.4]%, 200ms-delay = 85.6 [82.9, 89.2]%, effect size (no-delay - 200ms-delay) Hedges' g=-1.12; slope parameter: no-delay = 7.69 [5.82, 10.7] a.u., 200ms-delay = 6.61 [4.63, 8.74] a.u., Hedges' g=0.264; $t_{peak}$: no-delay = 494 [436, 557] ms, 200ms-delay = 552 [476, 680] ms, Hedges' g=-1.49), or on the STM-dependent stage (Fig. 5DE, $a_{peak}$: no-delay, median [C.I.] = 80.5 [75.6, 85.5]%, 200ms-delay = 80.9 [75.6, 84.9]%, Hedges' g=-0.147; $t_{decay}$: no-delay = 976 [332, 1642] ms, 200ms-delay = 580 [319, 1585] ms, Hedges' g=0.877; $t_{chance}$: no-delay = 2214 [1865, 2400] ms, 200ms-delay = 2400 [1935, 2400] ms, Hedges' g=-1.22).

Taken together, our results from varying the stimulus onset delay show that changes in RT distribution (and overall decision accuracy) are not necessarily accompanied by changes in the conditional accuracy function. The observed trend of decreased accuracy was accounted for by the fact that with a delay, there were more responses initiated before the sensory encoding was complete, or even before the stimulus was presented (i.e., 'impulsive' responses) (Fig.5B, histograms). To quantify such impulsivity, we propose an 'impulsivity index' (ImpI): ImpI = 1 – average (duration for which mice withhold responses /duration of the delay). Higher positive values of this index indicate greater impulsivity, with ImpI=1 indicating a complete inability to withhold responding in the face of stimulus delays ('maximally' impulsive). In the case of our mice, ImpI is ~0.6.

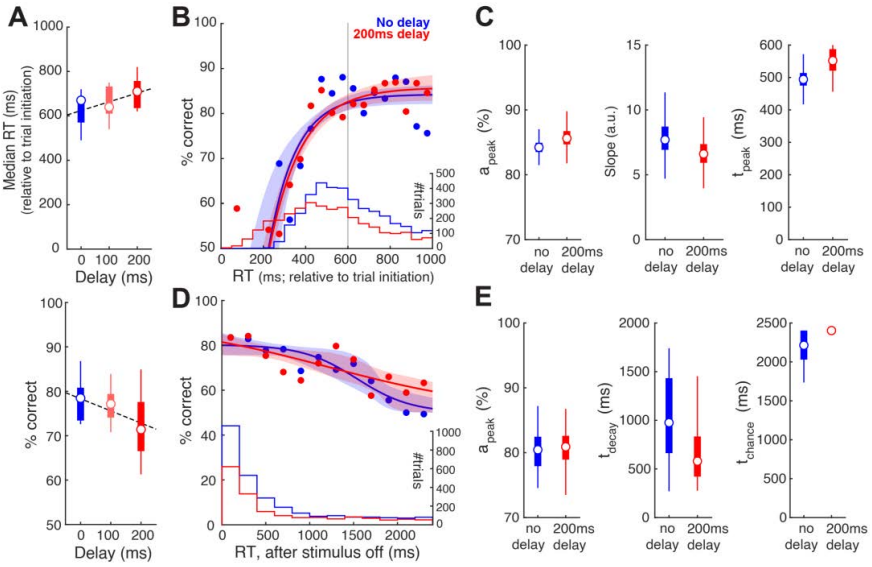


**Figure 5. Stimulus onset delay modulates RT distribution but not the conditional accuracy function. (A)** *Upper:* Plot of median RT, measured relative to train initiation, against stimulus onset delay (n=9 mice; p=0.094, 1-way ANOVA; effect size $\eta^2$=0.179; Pearson's correlation=0.422, p=0.028). Dashed line: Linear regression on RTs. *Lower:* Plot of response accuracy against stimulus onset delay (p=0.182, 1-way ANOVA; effect size $\eta^2$=0.132; Pearson's correlation=-0.358, p=0.067). **(B)** Conditional accuracy functions of the sensory encoding stage; Blue: trials with no delay; red: trials with 200ms delay; shaded bands: bootstrap confidence intervals (95%); confidence intervals overlap for the two datasets. Histograms: RT distributions. Grey vertical line: stimulus offset. **(C)** Key parameters of the CAF (sensory encoding stage) for trials with no delay vs. trials with 200ms delay. Box plots show the distribution of the bootstrapped estimates. **(D)** Conditional accuracy functions of the STM-dependent stage. Conventions as in B. (E) Key parameters of the CAF (STM-dependent stage) for trials with no delay vs. trials with 200ms delay. Conventions as in C.
**See also Fig. 5-1.**

## DISCUSSION

In this study, we quantify two distinct stages in the temporal dynamics of visual perceptual decisions in mice. First, a sensory encoding stage that is subject to the speed-accuracy tradeoff, and then, a short-term memory dependent stage in which decision performance decays once the stimulus disappears. We also demonstrate that the conditional accuracy function and the RT distribution can be affected independently by experimental manipulations. Whereas stimulus size, luminance and presence of a foil modulate the conditional accuracy function with minimal changes to the RT distribution, stimulus onset asynchrony modulates the RT distribution without changes to the conditional accuracy function. Additionally, our results yield numerical estimates of fundamental psychophysical constants of visual perceptual decision-making in mice. Taken together, this study establishes a quantitative platform for future work dissecting neural circuit underpinnings of the dynamics of visually guided decision-making in mice.

### Estimates of time constants of the dynamics of visual perceptual decision-making in mice

Our results yielded numerical estimates of the duration of sensory encoding (i.e., the window of temporal integration) as 200-320 ms across stimulus size and luminance in mice (Fig. 2). This estimate is similar to that in humans: the internal representation of a visual stimulus is thought to be constructed within the first 200-300 ms of stimulus presentation (Shibuya and Bundesen 1988, Busey and Loftus 1994, Vogel, Woodman et al. 2006, Bays, Gorgoraptis et al. 2011). On the other hand, we also obtained an estimate of

the duration of STM as 1700 ms. This constituted the period starting from stimulus offset to the last instant at which responses that are better than chance were *initiated* (Fig. 3D; $t_{chance}$ - $t_{delay}$ = ~1700 ms). This duration does not necessarily represent just the maintenance of visual stimulus information in STM, it could also represent maintenance of information about the motor response associated with the stimulus (and likely, a combination of the two). Notably, our estimate of the duration of viability of the labile internal representation in mice falls in the same range as has been reported from human studies (Sperling 1960, Posner and Keele 1967, Phillips and Baddeley 1971).

We have interpreted the decay in performance following stimulus offset as being due to loss of information in STM. A potential confounding factor to this interpretation is differences in the internal state of the animal – in selective attention, or more generally, task engagement. It is possible, for instance, that all the trials with longer RTs represent those in which mice did not pay attention to the stimulus (or more generally, were disengaged from the task), thereby being associated with lower accuracy. We believe this unlikely because attention/engagement was not varied systematically, here (unlike in the flanker task, Fig. 4). Even if loss of attention or engagement were a factor, any improvements in conditional accuracy due to increased attentiveness or engagement would only lengthen STM. From this perspective, our estimate of 1700 ms serves as a *lower bound* for the duration of STM.

This estimate of 1700 ms also represents a lower bound for working memory (WM). Whereas STM refers to the retention of information even when it is not accessible from the environment, WM refers additionally to the ability to manipulate this information and protect it from interference (Cowan 2008, Postle and Pasternak 2009). WM can be lengthened with training. For instance, in tasks that require animals hold information over an enforced delay period before responding, it has been reported that mice can learn to perform well with delay periods up to 5 sec (Liu, Gu et al. 2014). Here, by allowing the natural evolution of the dynamics of decision-making to occur without an imposed delay period, we were able to estimate the 'intrinsic' (lower bound for the) duration of STM.

**Estimates of the operating range of stimulus features for visual perceptual decision-making in mice**

This study also yielded estimates for the range of values of various stimulus features within which mice are able to discriminate the visual target. The smallest stimulus size and lowest luminance (tested) at which mice were able to discriminate orientation above chance were 25° and 2.00 cd/m$^2$, with mice performing at > 80% accuracy for most luminance values at that smallest size. The shortest stimulus that mice are able to discriminate above chance was ≤ 100ms (Fig. 3A). Further, based on the x-intercept of the CAF in sensory encoding stage (median [C.I.] = 236 [215, 253] ms, pooling all trials of various sizes and luminance from Fig. 2), we were able to refine this estimate to be ≤ 53ms (conservatively, after subtracting $t_{delay}$ = ~200 ms). This is consistent to the estimation (40-80 ms) from a previous study based on visual cortical activity (Resulaj, Ruediger et al. 2018). In a subgroup of animals (n=3), we tested if mice are able to discriminate orientation of the target stimulus (25°, 0.1cpd, 16.2 cd/m$^2$) when it was 50 ms long. Two out of the three mice showed a response accuracy higher than chance (accuracy = 57.9%, 210 correct out of 363 trials, p=0.002, binomial test; and 55.6%, 143/257, p=0.040, respectively), consistent with this refined estimate. These findings that mice are able to discriminate visual stimuli in demanding sensory contexts suggest that the visual perceptual abilities of mice may be underrated.

The best discrimination performance reported in mice (accuracies > 90%) have typically been obtained using large, often full-field, grating stimuli (Andermann, Kerlin et al. 2010, Long, Jiang et al. 2015). In our single target discrimination task, the best performance ranged lower, between 75-90% (Fig. 1C), consistent with our use of 'small' stimuli (relative to those typically used in mouse vision studies (Prusky, West et al. 2000, Prusky and Douglas 2004, Wong and Brown 2006, Busse, Ayaz et al. 2011, Long, Jiang et al. 2015)) and the lower visual acuity of mice. Indeed, in our pilot study, the performance plateaued at ~93% for a stimulus size ≥ 45° (Fig. 1-1CD). These results suggest that full-field stimuli may be effectively replaced by 45° stimuli to achieve best performance levels.

The best discrimination performance exhibited a dip at the highest luminance (Fig. 1C). This is potentially well accounted for by signal saturation: because the visual system adapts to the relevant range of stimulus luminance for best encoding (Ohzawa, Sclar et al. 1982), the interleaved presentation of stimuli with different luminance can render the maximum-luminance stimulus unfavorable because of signal saturation (Long, Jiang et al. 2015). Consistent with this idea, when the maximum-luminance stimulus (25°, 0.1cpd, 130 cd/m$^2$) was presented *alone* in block design (Fig. 1-1C, the green box at the left most, group median [C.I.] = 85.7 [77.6, 92.1] %), response accuracy was nominally higher than when it was presented interleaved with stimuli of varying luminance (Fig. 1C, the red box at the right most, group median [C.I.] = 79.7 [61.9, 91.9] %). These results indicate that a good upper bound for stimulus luminance in mouse experiments may be ~34 cd/m$^2$.

**Stimulus and task parameters modulate perceptual performance through a variety of mechanisms**

Increase in stimulus size and luminance both improved the overall discrimination performance of mice (Fig. 1). However, analysis of conditional accuracy revealed that whereas increasing each increased the peak conditional accuracy ($a_{peak}$), only increasing the stimulus luminance increased the slope of the CAF and resulted in a shorter $t_{peak}$ (Fig. 2). We propose that these differences in the CAF reflect differential mechanisms at one or more levels of the underlying sensory processing. Specifically, in our experiments, varying stimulus luminance (through varying the intensity of just the bright phase of the grating) also varied stimulus contrast (relative to the dark background). On the other hand, increasing stimulus size increased the total luminance without affecting contrast. Consequently, differential activation of lateral inhibitory mechanisms for spatial contrast may account for the observed differences in CAFs. Separately, whereas increasing stimulus size and luminance both increase the total number of photons impinging on the retina, increasing stimulus size would activate a broader spatial distribution of photoreceptors (at a fixed signal-to-noise ratio), while increasing stimulus luminance would cause a largely fixed group of photoreceptors to receive a higher density of photons (and higher signal-to-noise ratio). Consequently, differential mechanisms of sensory integration (of the two) may also account for the observed differences in CAFs.

Separately, manipulating attention (by presenting a flanker) and manipulating the stimulus onset asynchrony both caused a reduction in response accuracy (Fig.4B, 5A-lower panel). However, again, the analysis of conditional accuracy suggests that the mechanisms underlying the two are different: the capture of attention by the flanker interferes with the target's sensory encoding, whereas adding a pre-stimulus onset delay results in change of the RT distribution without affecting the CAF.

Taken together, our results demonstrate that although manipulating stimulus parameters or experimental conditions may induce seemingly similar changes in perceptual performance (overall accuracy), their underlying mechanisms could be different. The conditional accuracy analysis serves as an informative tool to explore these mechanisms and to understand the dynamics of perceptual decision making.

**Qualitative differences between stimulus features as well as between task-difficulties**

Across the various tasks and stimulus conditions that we studied here in mice, the sensory encoding stage ended rapidly, around 300ms. However, in a recent study in which rats discriminated the direction of motion of a patch of random dots, the sensory encoding stage continued through at least 1.5 s (the longest RT bin reported (Shevinsky and Reinagel 2019)). We propose that this difference in the duration of sensory encoding may be due to fundamentally different nature of stimulus features used in these two studies. Consistent with this proposal, a study on human visual psychophysics (Burr and Santoro 2001) has reported a temporal integration window of 200-300ms when stimulus contrast of a patch of random dots was varied (similar to our results in mice), but a substantially longer integration window of 3s when their motion coherence levels were varied (similar to the above-mentioned results in rats).

Separately, our results also highlight that 'task difficulty' may be altered in qualitatively different ways, producing distinct outcomes on behavior. In the literature, task difficulty is often increased by making target stimuli more ambiguous or by introducing distracters (which we did also). Such manipulations often cause subjects (animals) to respond slower, allowing them time to gather more information to produce better performance (which we found, as well). However, when we shortened stimulus duration, which can plausibly be considered to also increase task difficulty, we found the opposite result – mice responded faster as the target stimulus became shorter (Fig. 3B). This potentially counter-intuitive effect (faster RTs for a 'more difficult' task) is explained well by the conditional accuracy analysis (Fig. 3C). Whereas shortening the stimulus duration makes the task more difficult, responding more slowly to shorter stimuli does not grant a perceptual benefit to the animals: once the stimulus has disappeared, withholding responses for longer would only increase the risk of losing information owing to memory decay. In other words, short stimuli impose a 'time pressure' on animals to make decisions quickly. Thus, task difficulty may be altered in qualitatively different ways, with distinct behavioral effects.

**Optimal sensory sampling during visual perceptual decision-making in mice**

An intriguing observation in our study is that across tasks, the peak of RT distribution (the RT bin with the largest number of trials) always seemed to occur around $t_{peak}$ (Fig. 2AC, 4C). Since the RT distribution can vary independently of the conditional accuracy function (as demonstrated in Fig. 5), there is no *priori* reason that the peak of RT distribution and the $t_{peak}$ must change together. We propose that responding with RTs close to $t_{peak}$ may be an optimal behavioral strategy for the mice. As indicated by the conditional accuracy function, mouse response accuracy increased as RT increased until it reached a plateau at $t_{peak}$. Responding earlier than $t_{peak,}$ therefore, would sacrifice accuracy, while responding later than $t_{peak}$ would needlessly delay response (reducing the reward rate). Consequently, responding with the peak of RT distribution being equal to $t_{peak}$ would be optimal. Testing this optimality hypothesis would require future experiments to manipulate the temporal integration window ($t_{peak}$) substantially (much more than the 40 ms - 120 ms change we find in Figs. 2BD, 4D) – for instance, by manipulating stimulus coherence (Burr and Santoro 2001) or the volatility of environment (Piet, El Hady et al. 2018), and to ask if this is accompanied by a commensurate shift in peak RT.

**EXTENDED DATA**

Extended data (Fig. 1-1, 2-1, 2-2, 3-1, and 5-1) and legends are included.

**REFERENCES**

Andermann, M. L., A. M. Kerlin and R. C. Reid (2010). "Chronic cellular imaging of mouse visual cortex during operant behavior and passive viewing." Front Cell Neurosci **4**: 3.

Barrouillet, P. and V. Camos (2012). "As time goes by: Temporal constraints in working memory." Current Directions in Psychological Science **21**(6): 413-419.

Bays, P. M., N. Gorgoraptis, N. Wee, L. Marshall and M. Husain (2011). "Temporal dynamics of encoding, storage, and reallocation of visual working memory." Journal of vision **11**(10): 6-6.

Brown, J. (1958). "Some tests of the decay theory of immediate memory." Quarterly Journal of Experimental Psychology **10**(1): 12-21.

Burgess, C. P., A. Lak, N. A. Steinmetz, P. Zatka-Haas, C. Bai Reddy, E. A. K. Jacobs, J. F. Linden, J. J. Paton, A. Ranson, S. Schroder, S. Soares, M. J. Wells, L. E. Wool, K. D. Harris and M. Carandini (2017). "High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice." Cell Rep **20**(10): 2513-2524.

Burr, D. C. and L. Santoro (2001). "Temporal integration of optic flow, measured by contrast and coherence thresholds." Vision research **41**(15): 1891-1899.

Busey, T. A. and G. R. Loftus (1994). "Sensory and cognitive components of visual information acquisition." Psychological Review **101**(3): 446.

Busse, L., A. Ayaz, N. T. Dhruv, S. Katzner, A. B. Saleem, M. L. Scholvinck, A. D. Zaharia and M. Carandini (2011). "The detection of visual contrast in the behaving mouse." J Neurosci **31**(31): 11351-11361.

Carandini, M. and A. K. Churchland (2013). "Probing perceptual decisions in rodents." Nature neuroscience **16**(7): 824.

Coltheart, M. (1980). "Iconic memory and visible persistence." Perception & psychophysics **27**(3): 183-228.

Cowan, N. (2008). "What are the differences between long-term, short-term, and working memory?" Progress in brain research **169**: 323-338.

Dick, A. (1974). "Iconic memory and its relation to perceptual processing and other memory mechanisms." Perception & Psychophysics **16**(3): 575-596.

Eriksen, B. A. and C. W. Eriksen (1974). "Effects of noise letters upon the identification of a target letter in a nonsearch task." Perception & Psychophysics **16**(1): 143-149.

Fan, J., B. D. McCandliss, T. Sommer, A. Raz and M. I. Posner (2002). "Testing the efficiency and independence of attentional networks." J Cogn Neurosci **14**(3): 340-347.

Glickfeld, L. L., M. H. Histed and J. H. Maunsell (2013). "Mouse primary visual cortex is used to detect both orientation and contrast changes." J Neurosci **33**(50): 19416-19422.

Glickfeld, L. L., R. C. Reid and M. L. Andermann (2014). "A mouse model of higher visual cortical function." Current opinion in neurobiology **24**: 28-33.

Gold, J. M., R. F. Murray, A. B. Sekuler, P. J. Bennett and R. Sekuler (2005). "Visual memory decay is deterministic." Psychological Science **16**(10): 769-774.

Guo, Z. V., S. A. Hires, N. Li, D. H. O'Connor, T. Komiyama, E. Ophir, D. Huber, C. Bonardi, K. Morandell, D. Gutnisky, S. Peron, N. L. Xu, J. Cox and K. Svoboda (2014). "Procedures for behavioral experiments in head-fixed mice." PLoS One **9**(2): e88678.

Heitz, R. P. (2014). "The speed-accuracy tradeoff: history, physiology, methodology, and behavior." Frontiers in neuroscience **8**: 150.

Hentschke, H. and M. C. Stüttgen (2011). "Computation of measures of effect size for neuroscience data sets." European Journal of Neuroscience **34**(12): 1887-1894.

Histed, M. H., L. A. Carvalho and J. H. Maunsell (2012). "Psychophysical measurement of contrast sensitivity in the behaving mouse." J Neurophysiol **107**(3): 758-765.

Histed, M. H. and J. H. Maunsell (2014). "Cortical neural populations can guide behavior by integrating inputs linearly, independent of synchrony." Proceedings of the National Academy of Sciences **111**(1): E178-E187.

Huberman, A. D. and C. M. Niell (2011). "What can mice tell us about how vision works?" Trends Neurosci **34**(9): 464-473.

Liu, D., X. Gu, J. Zhu, X. Zhang, Z. Han, W. Yan, Q. Cheng, J. Hao, H. Fan, R. Hou, Z. Chen, Y. Chen and C. T. Li (2014). "Medial prefrontal activity during delay period contributes to learning of a working memory task." Science **346**(6208): 458-463.

Long, M., W. Jiang, D. Liu and H. Yao (2015). "Contrast-dependent orientation discrimination in the mouse." Sci Rep **5**: 15830.

Mar, A. C., A. E. Horner, S. R. Nilsson, J. Alsio, B. A. Kent, C. H. Kim, A. Holmes, L. M. Saksida and T. J. Bussey (2013). "The touchscreen operant platform for assessing executive function in rats and mice." Nat Protoc **8**(10): 1985-2005.

McElree, B. and B. A. Dosher (1989). "Serial position and set size in short-term memory: the time course of recognition." Journal of Experimental Psychology: General **118**(4): 346.

Meier, P., E. Flister and P. Reinagel (2011). "Collinear features impair visual detection by rats." Journal of vision **11**(3): 22-22.

Miller, J. (1991). "The flanker compatibility effect as a function of visual angle, attentional focus, visual transients, and perceptual load: a search for boundary conditions." Percept Psychophys **49**(3): 270-288.

Nomura, Y., S. Ikuta, S. Yokota, J. Mita, M. Oikawa, H. Matsushima, A. Amano, K. Shimonomura, Y. Seya and C. Koike (2019). "Evaluation of critical flicker-fusion frequency measurement methods using a touchscreen-based visual temporal discrimination task in the behaving mouse." Neuroscience research **148**: 28-33.

Ohzawa, I., G. Sclar and R. D. Freeman (1982). "Contrast gain control in the cat visual cortex." Nature **298**(5871): 266-268.

Phillips, W. and A. Baddeley (1971). "Reaction time and short-term visual memory." Psychonomic Science **22**(2): 73-74.

Piet, A. T., A. El Hady and C. D. Brody (2018). "Rats adopt the optimal timescale for evidence integration in a dynamic environment." Nature communications **9**(1): 1-12.

Posner, M. I. and S. W. Keele (1967). "Decay of visual information from a single letter." Science **158**(3797): 137-139.

825 Postle, B. and T. Pasternak (2009). "Short term and working memory." Encyclopedia of Neuroscience. San Diego,
826 CA: Elsevier: 783-789.
827 Prusky, G. T. and R. M. Douglas (2004). "Characterization of mouse cortical spatial vision." Vision Res **44**(28):
828 3411-3418.
829 Prusky, G. T., P. W. West and R. M. Douglas (2000). "Behavioral assessment of visual acuity in mice and rats."
830 Vision Res **40**(16): 2201-2209.
831 Ratcliff, R. (1978). "A theory of memory retrieval." Psychological Review **85**(2): 59-108.
832 Ratcliff, R., P. L. Smith, S. D. Brown and G. McKoon (2016). "Diffusion Decision Model: Current Issues and
833 History." Trends Cogn Sci **20**(4): 260-281.
834 Resulaj, A., S. Ruediger, S. R. Olsen and M. Scanziani (2018). "First spikes in visual cortex enable perceptual
835 discrimination." Elife **7**: e34044.
836 Ricker, T. J., E. Vergauwe and N. Cowan (2016). "Decay theory of immediate memory: From Brown (1958) to
837 today (2014)." The Quarterly Journal of Experimental Psychology **69**(10): 1969-1995.
838 Seabrook, T. A., T. J. Burbridge, M. C. Crair and A. D. Huberman (2017). "Architecture, function, and assembly of
839 the mouse visual system." Annual review of neuroscience **40**: 499-538.
840 Shevinsky, C. A. and P. Reinagel (2019). "The interaction between elapsed time and decision accuracy differs
841 between humans and rats." Frontiers in neuroscience **13**: 1211.
842 Shibuya, H. and C. Bundesen (1988). "Visual selection from multielement displays: measuring and modeling effects
843 of exposure duration." Journal of Experimental Psychology: Human Perception and Performance **14**(4): 591.
844 Siegel, M., A. K. Engel and T. H. Donner (2011). "Cortical network dynamics of perceptual decision-making in the
845 human brain." Frontiers in human neuroscience **5**: 21.
846 Smith, P. L. and R. Ratcliff (2009). "An integrated theory of attention and decision making in visual signal
847 detection." Psychol Rev **116**(2): 283-317.
848 Speed, A., J. Del Rosario, N. Mikail and B. Haider (2020). "Spatial attention enhances network, cellular and
849 subthreshold responses in mouse visual cortex." Nature communications **11**(1): 1-11.
850 Sperling, G. (1960). "The information available in brief visual presentations." Psychological monographs: General
851 and applied **74**(11): 1.
852 Stanford, T. R., S. Shankar, D. P. Massoglia, M. G. Costello and E. Salinas (2010). "Perceptual decision making in
853 less than 30 milliseconds." Nature neuroscience **13**(3): 379.
854 Steinemann, N. A., R. G. O'Connell and S. P. Kelly (2018). "Decisions are expedited through multiple neural
855 adjustments spanning the sensorimotor hierarchy." Nature communications **9**(1): 1-13.
856 Thura, D. and P. Cisek (2014). "Deliberation and commitment in the premotor and primary motor cortex during
857 dynamic decision making." Neuron **81**(6): 1401-1416.
858 Uchida, N., A. Kepecs and Z. F. Mainen (2006). "Seeing at a glance, smelling in a whiff: rapid forms of perceptual
859 decision making." Nature Reviews Neuroscience **7**(6): 485-491.
860 Umino, Y., R. Pasquale and E. Solessio (2018). "Visual temporal contrast sensitivity in the behaving mouse shares
861 fundamental properties with human psychophysics." eNeuro **5**(4).
862 Vogel, E. K., G. F. Woodman and S. J. Luck (2006). "The time course of consolidation in visual working memory."
863 Journal of Experimental Psychology: Human Perception and Performance **32**(6): 1436.
864 Voss, A., M. Nagler and V. Lerche (2013). "Diffusion models in experimental psychology: a practical introduction."
865 Exp Psychol **60**(6): 385-402.
866 Voss, A., J. Voss and V. Lerche (2015). "Assessing cognitive processes with diffusion model analyses: a tutorial
867 based on fast-dm-30." Front Psychol **6**: 336.
868 Wang, L. and R. J. Krauzlis (2018). "Visual Selective Attention in Mice." Curr Biol **28**(5): 676-685.e674.
869 Whitney, D. and D. M. Levi (2011). "Visual crowding: a fundamental limit on conscious perception and object
870 recognition." Trends Cogn Sci **15**(4): 160-168.
871 Wickelgren, W. A. (1977). "Speed-accuracy tradeoff and information processing dynamics." Acta psychologica
872 **41**(1): 67-85.
873 Wilming, N., P. R. Murphy, F. Meyniel and T. H. Donner (2020). "Large-scale dynamics of perceptual decision
874 information across human cortex." Nature communications **11**(1): 1-14.
875 Wong, A. A. and R. E. Brown (2006). "Visual detection, pattern discrimination and visual acuity in 14 strains of
876 mice." Genes Brain Behav **5**(5): 389-403.
877 Yang, Y., M. DeWeese, G. Otazu and A. Zador (2008). "Millisecond-scale differences in neural activity in auditory
878 cortex can drive decisions." Nature Precedings: 1-1.
879 You, W.-K. and S. P. Mysore (2020). "Endogenous and exogenous control of visuospatial selective attention in
880 freely behaving mice." Nature Communications **11**(1): 1-14.

881 Zariwala, H. A., A. Kepecs, N. Uchida, J. Hirokawa and Z. F. Mainen (2013). "The limits of deliberation in a
882 perceptual decision task." Neuron **78**(2): 339-351.
883 Zhang, W. and S. J. Luck (2009). "Sudden death and gradual decay in visual working memory." Psychological
884 science **20**(4): 423-428.

885

886

# Extended Data

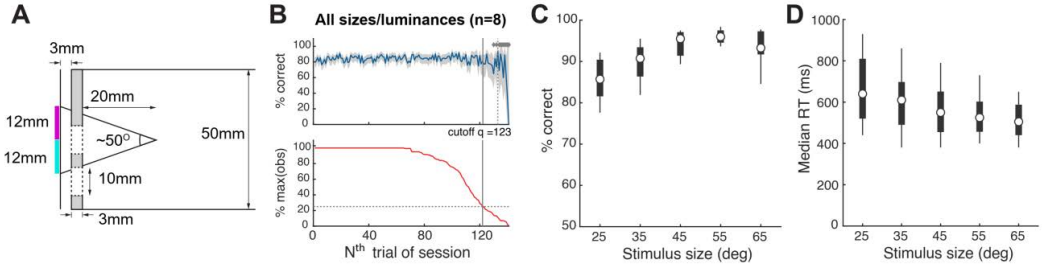


**Figure 1-1. Extended data for Figure 1.**

**(A)** Lateral view of the schematic experimental setup showing the relative position of the touchscreen (leftmost vertical line), the plexiglass mask (grey-filled vertical bar), and the tube within which mice move (50 mm diameter); the plexiglass mask is positioned 3 mm in front of the touchscreen. Dashed lines indicate the central response hole (lower dashed lines), and left/right response holes (upper dashed lines; 10 mm diameter). For single-stimulus discrimination, the center of the stimulus is aligned with the center of left/right response holes in elevation, and with the central hole in azimuth (see Fig.1A). For experiments involving two stimulus locations (i.e., flanker task), the upper (magenta) and lower (cyan) locations of the stimulus are indicated as colored bars (see also Fig. 4A). The 60 pixels x 60 pixels (12mm x 12mm) stimulus subtends a visual angle of 25º when viewed from 20 mm front of the plexiglass mask. **(B)** Identification of trials towards the end of the 30 min behavioral sessions that corresponded to animals being poorly engaged in the task (Methods and (You and Mysore 2020)). Top panel: Time course of overall response accuracy across mice as a function of trial number within sessions. Accuracy obtained from trials pooled across all mice and sessions, and computed as a function of trial number within session (blue; Methods). Grey shading: bootstrapped estimates of the 95% confidence interval of the accuracy (gray; Methods). Diamonds on top: trials whose accuracy not significantly different from chance. Dashed vertical line: first trial at which the accuracy was not different from chance (50%), and stayed indistinguishable from chance for 3/5 of the next 5 trials (Methods). Data show increased variability and worse performance towards the end of sessions. Bottom panel: Number of actual observations across mice for each trial number, as a percentage of the maximal number of possible observations (Σ mice*sessions), plotted as a function of trial number within session (red). Solid vertical line: first trial at which the number of observations drops below 25%. Data show drop in the number of observations available to reliably assess performance towards the end of sessions. Based on these data, all trials above 122 of each behavioral session of this experiment were dropped from analysis (Methods). Results in Fig. 1 are based on data from trials 1-122 from each behavioral session. (**C**) Response accuracy as a function of stimulus size (n=9 mice; p=0.001, 1-way ANOVA). In these experiments, stimulus size was manipulated independently (without manipulation of luminance; unlike in Figure 1). All stimuli were at the highest luminance (130 cd/m$^2$). **(D**) Median RT as a function of stimulus size (n=9 mice; p=0.205, 1-way ANOVA).

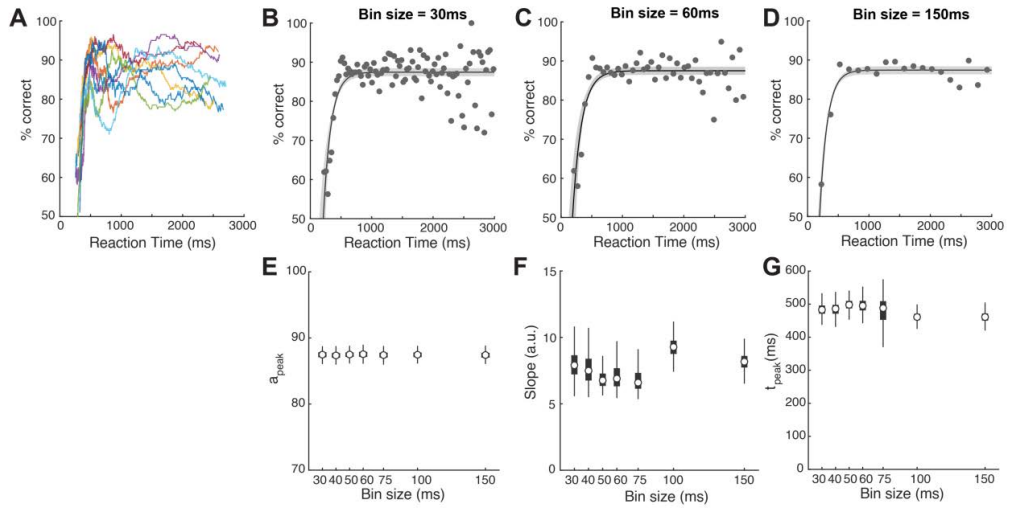


**Figure 2-1. Extended data for Figure 2: CAFs of individual mice and effect of bin size.**
**(A)** The general pattern of conditional accuracy curves across mice. Each color represents one single mouse. Each curve was generated by pooling all trials (of various stimulus size and luminance) from one mouse, sort the trials by RT, and then do a moving average (window size = 200 trials) to plot the mean accuracy (y) at mean RT (x) of the time window. **(B-D)** Fitting of the conditional accuracy function (CAF) in various bin sizes. (B) Bin size = 30ms; (C) Bin size = 60ms; (D) Bin size = 150ms; **(E-G)** Estimates of the quantitative metrics of the CAF in various bin sizes. (E) peak conditional accuracy ($a_{peak}$); (F) slope parameter; and (G) time to reach peak conditional accuracy ($t_{peak}$).

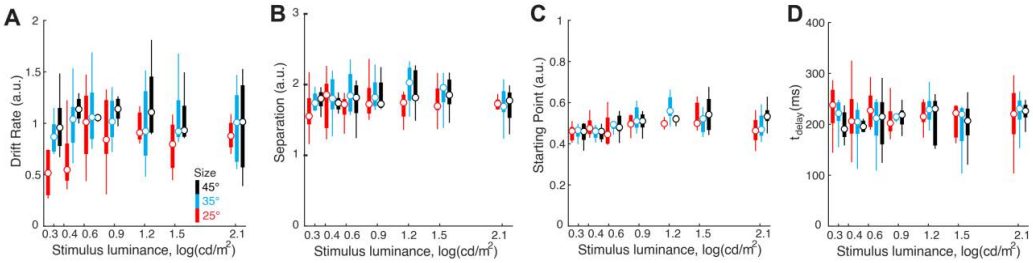


**Figure 2-2. Extended data for Figure 2: Estimates of all four parameters of the drift diffusion model. (A)** Drift rate; 2-way ANOVA, p=0.028 (luminance), p<0.001 (size), p=0.767 (interaction). **(B)** Boundary separation; 2-way ANOVA, p=0.171 (luminance), p=0.026 (size), p=0.953 (interaction). **(C)** Starting point; 2-way ANOVA, p<0.001 (luminance), p=0.325 (size), p=0.098 (interaction). **(D)** $t_{delay}$; 2-way ANOVA, p=0.523 (luminance), p=0.308 (size), p=0.931 (interaction).

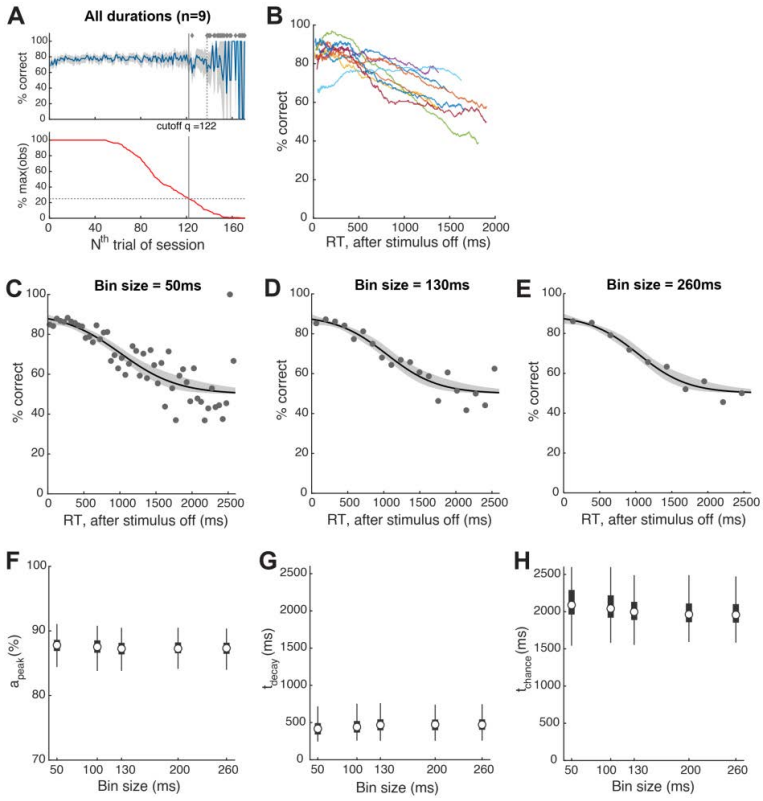


**Figure 3-1. Extended data for Figure 3: CAFs of individual mice and effect of bin size. (A)** Identification of trials towards the end of the 30 min behavioral sessions that corresponded to animals being poorly engaged in the task (Methods); conventions identical to those in Fig.1-1B. **(B)** The general pattern of conditional accuracy curves across mice. Each color represents one single mouse. Each curve was generated by pooling all trials (of various stimulus size and luminance) from one mouse, sort the trials by RT, and then do a moving average (window size = 200 trials) to plot the mean accuracy (y) at mean RT (x) of the time window. **(C-E)** Fitting of the conditional accuracy function (CAF) in various bin sizes. (C) Bin size = 50ms; (D) Bin size = 130ms; (E) Bin size = 260ms; **(F-H)** Estimates of the quantitative metrics of the CAF in various bin sizes. (F) peak conditional accuracy ($a_{peak}$); (G) the time at which conditional accuracy started to decay ($t_{decay}$); and (G) the time at which conditional accuracy fell to the chance level ($t_{chance}$).

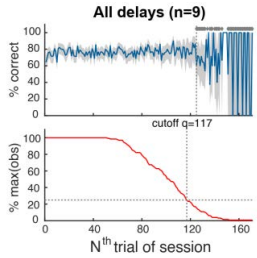


**Figure 5-1. Extended data for stimulus onset delay experiment.**
Identification of trials towards the end of the 30 min behavioral sessions that corresponded to animals being poorly engaged in the task (Methods). All conventions are as in Fig.1-1B. Based on these data, all trials above 116 of each behavioral session of this experiment were dropped from analysis. Results in Fig.5 are based on data from trials 1-116 from each behavioral session.