Visual psychophysics and limits of visual discrimination performance in freely behaving mice

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

*Corresponding author: shreesh.mysore@jhu.edu

ABSTRACT

Mice are being used increasing commonly to study visually guided behaviors. To help frame the design of visual tasks in mice, we explored limits of mouse visual behavior using a touchscreen-based 2AFC orientation discrimination task in unrestrained animals. We found that mice were able to discriminate targets as small as 25°, as brief as 100 ms, and with an 'impulsivity index' of 0.6. They were able to perform well a rudimentary visual search task, exhibiting classic psychometric curves to the relative contrast between target and foil. Using a combination of conditional accuracy analysis and drift diffusion modeling, we estimated the time for sensory encoding in mice as 300 ms, and the duration of their visual short-term memory as 1700 ms. Our results reveal surprising parallels between aspects of mouse and human visual behavior, and suggest that visual perceptual abilities of mice may be underappreciated.

INTRODUCTION

Recent years have seen a rise in the use of the laboratory mouse for the study of the visual system [1-3] and visually guided behaviors [4-10]. This has been driven partly by the wealth of modern, genetics-based tools available for neural interrogation in mice. Nonetheless, because of their lower visual acuity than primates[1, 10-12]), as well as their perceived impulsivity [13, 14], there have been concerns that mice may not be ideally suited to study higher visual cognitive function [1]. These concerns are now being somewhat alleviated, for instance, by the successful demonstration of the study of primate-like visuospatial selective attention in both head-fixed [15] as well as feely behaving mice [16]. In light of these developments, an understanding of the limits of mouse visual performance is imperative for the appropriate design of visual tasks in mice, but represents a gap in our knowledge. Whereas psychophysical curves to stimulus contrast as well as spatial frequency have been obtained in mice through a variety of methods ([8-12]), an in-depth exploration of the operating range of task features (presence and properties of competing foil), other stimulus features (size, duration, etc.), as well as of key perceptual processes (duration of visual short term memory, window of sensory integration, etc) underlying mouse visual behavior is lacking. Because stimulus feature discrimination is a core module in studies of visually guided behavior, here, we explored the limits of mouse visual performance using a 2AFC orientation discrimination task as the basis. Considering the highly exploratory nature of the native behavior of mice, we examined visual performance limits in unrestrained, freely behaving mice, and did so using a touchscreen-based set up [16, 17].

In a series of experiments, we examined the effect of stimulus size, contrast, duration, and delay, as well as the presence, relative information content and relative contrast of a foil, on mouse performance. We used standard behavioral metrics of response accuracy, reaction time (RT), perceptual sensitivity (d') and decision criterion, to quantify aspects of mouse performance. Our results not only revealed that mice performed successfully in these experiments despite being challenged progressively more with different manipulations, but also identified limiting values in the stimulus/task features for successful performance. Moreover, by applying the conditional accuracy analysis [18-21], we identified two distinct stages in the time-course of their behavior within a trial – a temporally limited sensory encoding stage [22-26] in which

response speed and response accuracy exhibit a tradeoff, and a second stage, impacted by visual short-term memory (VSTM; [27-35]), in which they do not. Combining these results with those from drift diffusion modeling of RT distributions allowed us to estimate the sensory encoding time of mice, the length of their visual short-term memory, the shortest visual stimulus that is informative, and the longest stimulus beyond which no additional benefit in response accuracy is seen. Finally, by varying stimulus onset delay, we quantitatively estimated impulsivity of mice via an 'impulsivity index'. Our results provide a window into the operational range of key parameters in mouse visual discrimination behavior, and can serve as a quantitative behavioral guide for future studies exploring the neural circuit basis of visual cognition in mice.

RESULTS

All the behavioral tasks in this study involved a touchscreen-based setup described previously [16, 17] (Methods). Briefly, freely behaving mice were placed in a plexiglass tube within a soundproof operant chamber equipped with a touch-sensitive screen, and a reward well located at the opposite face of the box from the touchscreen (Fig. 1A, S1A). A plexiglass sheet, with three holes corresponding to the locations at which the mouse was allowed to interact with the touchscreen by a nose-touch, was placed in front of it. All trials began with a nose-touch on a bright zeroing-cross presented within the lower central hole. Immediately following nose-touch, visual stimuli (bright objects on a dark background) were presented on the screen. The lateralized upper holes served as response ports for the animals to report their behavioral choice (left vs. right nose-touch). Behavioral data were collected from daily sessions that lasted 30 minutes for each mouse.

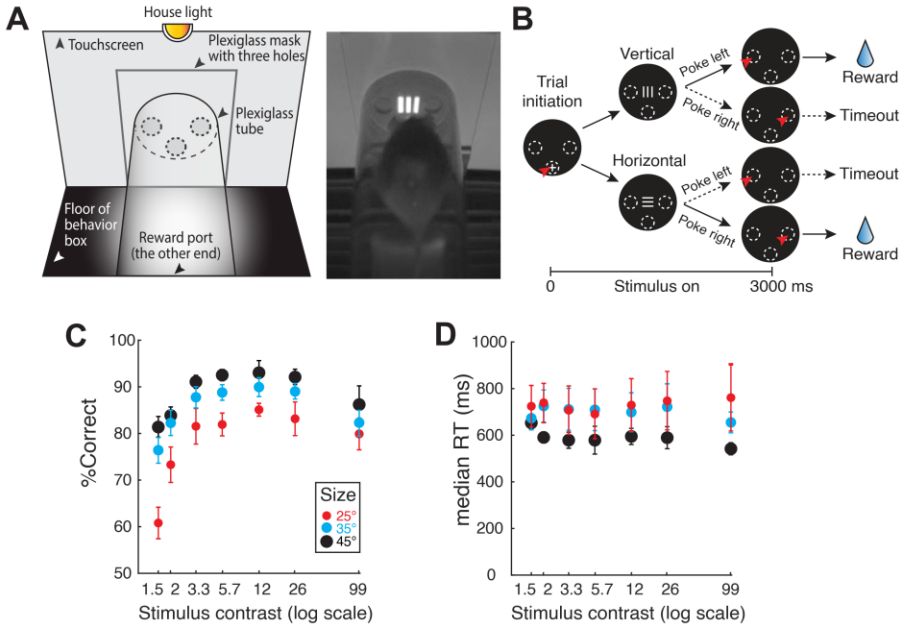

Figure 1. Stimulus contrast and size modulate orientation discrimination performance in freely behaving mice. (A) Left: Schematic of touchscreen-based experimental setup showing key components. Right: Snapshot of freely behaving mouse facing a visual stimulus on the touchscreen. **(B)** Schematic of 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 contrast (luminance$_{Bright}$ /

luminance$_{Dark}$; log scale; Methods). Different colors correspond to different target sizes. Data: mean ± s.e.m; n= 8 mice. 2-way ANOVA, p<0.001 (contrast), p<0.001 (size), p=0.498 (interaction). **(D)** Plot of median reaction time (RT) against stimulus contrast (log scale). 2-way ANOVA, p=0.99 (contrast), p=0.004 (size), p=1 (interaction). **See also Fig. S1.**

Stimulus contrast and size modulate mouse performance in discriminating grating orientation.

To explore the limits of visual discrimination of mice, we started by examining their performance on a single stimulus orientation discrimination task, in which we systematically varied the contrasts and sizes of the stimulus. Upon trial initiation, a grating stimulus ("target"), whose orientation could be either vertical or horizontal, was presented at the center of the screen for up to 3 seconds (Fig.1B; Methods). Mice were trained to respond to the orientation of the target with an appropriate nose-touch (vertical → left and horizontal → right). Mice were allowed to respond at any time during stimulus presentation, with stimulus presentation terminating automatically upon response. A correct response resulted in a beep (1s, 600Hz), followed by reward delivery (10uL water) at the port located at the opposite end of the chamber from the touchscreen. An incorrect response resulted in a 5-second pause, during which the house light was illuminated, following which the central cross became available once again for the mouse to initiate the next trial (Methods).

Three different sizes of the target were tested: 25º (60 x 60 pixels^2), 35º (84 x 84 pixels^2), and 45º (108 x 108 pixels^2), and for each size, seven different contrasts were tested (luminance$_{bright}$/luminance$_{dark}$ = 1.5, 2, 3.3, 5.7, 12, 26, 99; Methods). The spatial frequency of the grating was chosen to be 0.1 cycles/degree (24 pixels/cycle; for this task as well as all subsequent tasks), based on published reports that this value is within the range of spatial frequencies at which mice have the best visual contrast sensitivity [10, 11].

We found that both the stimulus size and contrast significantly modulated discrimination performance in mice (Fig. 1C, 2-way ANOVA, main effect of size, p<0.001; main effect of contrast, p<0.001; Fig. S1CD). Mice discriminated the orientation better, in general, when the target was of higher contrast, with performance plateauing at a contrast of 12 ("best" contrast) for all target sizes. This was reflected both in discrimination accuracy (Fig. 1C) as well as in perceptual sensitivity (Fig. S1BC; Methods); decision criterion was largely unaffected by stimulus contrast (Fig. S1D). Higher stimulus contrasts (than 12) did not provide an additional benefit for perceptual judgements (Fig. 1C and S1C). Notably, even at the lowest contrast tested (1.5), mice were able to discriminate target orientation better than chance (50%; Fig. 1C; red dot at the left lower corner, p=0.039, Wilcoxon signed rank test)

Along similar lines, mice discriminated target orientation better when the stimulus was larger, with discrimination accuracy plateauing at 93% correct (Fig. 1C) and perceptual discriminability at 3.37 (Fig. S1C) for a stimulus size of 45º (and best contrast); larger stimulus sizes did not provide additional benefits for perceptual judgements (Fig. S1F-H). There was no significant effect of target size on response criterion (Fig. S1D). Notably, for the smallest target size that we tested (25º), mice were still able to discriminate the orientation with >80% accuracy for most of the stimulus contrasts (≥3.3; Fig. 1C, red data).

Analysis of reaction times (RTs) revealed a significant effect of stimulus size - mice responded faster when the stimulus was larger, but there were no significant changes in median RT (surprisingly) with change in stimulus contrast (Fig.1D, two-way ANOVA; main effect of size, p=0.004; main effect of contrast, p=0.998; interaction, p=1).

Together, these results revealed a systematic effect of target contrast as well as size on discrimination accuracy, driven primarily by their effect on perceptual sensitivity rather than response criterion. Median RT negatively correlated with target size (Fig. S1G, Pearson's ρ=0.83, p=0.08), but exhibited no significant effect with respect to target contrast.

Stimulus size and contrast modulate the conditional accuracy function (CAF).
To investigate in greater detail the performance of mice on this task, we made use of the natural variability of the RT data and adopted the 'conditional accuracy analysis'[18-21]; Methods). This involves examining the dependence of mouse discrimination accuracy on RT, producing a conditional accuracy function (CAF; Fig. 2A). This analysis links the two commonly used metrics of behavioral performance, namely, accuracy and reaction time, with the overall accuracy being the dot product of the CAF with the RT distribution. As a result, application of the conditional analysis can help decompose observed change in response accuracy following any experimental manipulation, into changes in the CAF, in RTD, or both.

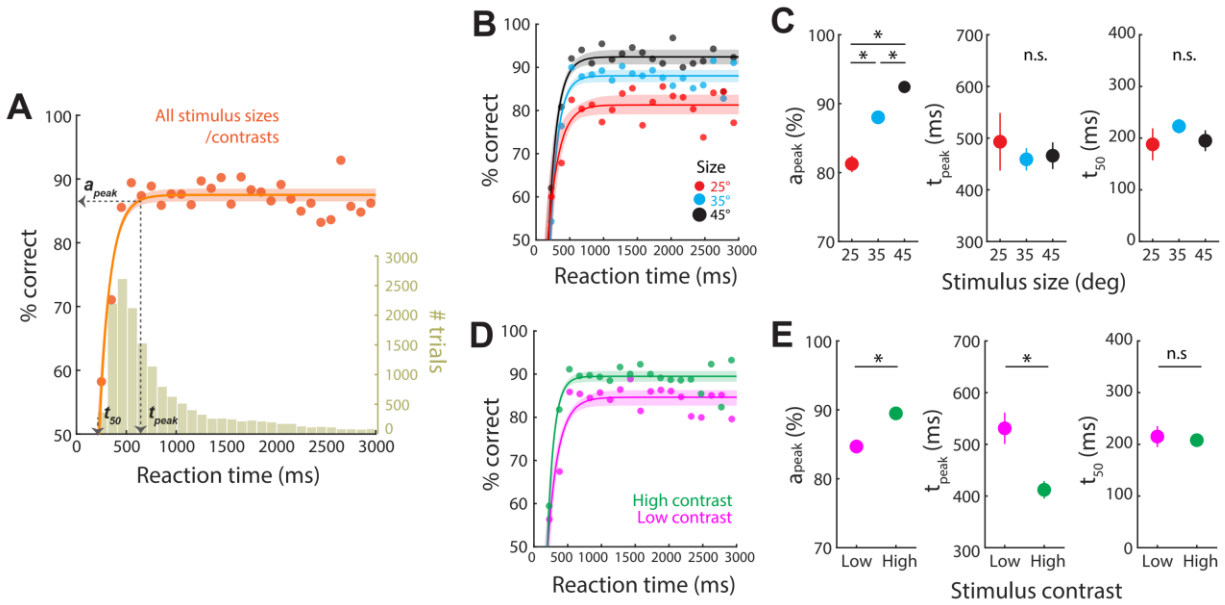

Figure 2. Stimulus contrast and size modulate the sensory encoding regime of the conditional accuracy function. (A) Plot of accuracy as a function of RT bins (conditional accuracy) using same dataset as Fig. 1. Orange dots: Data pooled across all stimulus sizes and contrasts, n=8 mice; RT bin size = 100 ms. Orange curve: Conditional accuracy function, CAF (best-fit rising asymptotic function; Methods); light orange shading: 95% CI of the fit (Methods). Indicated are three key parameters (a_{peak}, t_{peak}, and t_{50}) describing the sensory encoding stage of the CAF - the initial period during which accuracy improves for longer RT values, exhibiting a tradeoff between speed and accuracy (see text; Methods). Peak accuracy (a_{peak}): mean ± s.d.= 87.5 ± 0.5%; time to reach peak accuracy (t_{peak}): 462 ± 13 ms; time at which accuracy just exceeds 50% (t_{50}): 236 ± 10 ms. Gold histogram: RT distribution (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)** CAFs for targets of various sizes (black: 45º; blue: 35º; red: 25º); conventions as in A. **(C)** Plots of key CAF parameters for different target sizes. Left panel: a_{peak}; middle panel: t_{peak}; right panel: t_{50}. Data show mean ± s.t.d of distribution of bootstrapped estimates (Methods). '*' ('n.s.'): $p<0.05$ ($p>0.05$), paired permutation tests followed by HBMC correction (Methods). a_{peak}: $p<0.001$ (25 ° vs. 35°), $p<0.001$ (35 ° vs. 45°), $p<0.001$ (25 ° vs. 45°); t_{peak}: $p=0.398$ (25 ° vs. 35°), $p=0.827$ (35 ° vs. 45°), $p=0.576$ (25 ° vs. 45°); t_{50}: $p=0.226$ (25 ° vs. 35°), $p=0.127$ (35 ° vs. 45°), $p=0.918$ (25 ° vs. 45°). **(D)** CAFs for targets of different contrast conditions (magenta: 'low' contrast - first three contrast levels from Fig. 1C; green: 'high' contrast - last four contrast levels; Methods); conventions as in A. **(E)** Plots of key CAF parameters for different contrast conditions; conventions and statistical methods as in C. a_{peak}: $p<0.001$ (low vs. high contrast conditions); t_{peak}: $p<0.001$; t_{50}: $p=0.747$.

As a first step, we pooled trials from all mice (n=8) across the different trial conditions (3 sizes x 7 contrasts), sorted them based on RT, and calculated conditional accuracy for each RT bin (100ms;

Fig.2A; Methods). We found that there were two distinct regimes in the relationship between conditional accuracy and RT: (1) for responses with RT between 200 ms (minimal RT) and 500ms, conditional accuracy was better for longer RT (Pearson's ρ=0.99, p=0.02), consistent with the 'speed-accuracy tradeoff' (SAT, Heitz2014); and (2) for responses with RT>500ms (and up to 3s, the task limit), conditional accuracy was independent of RT (Pearson's ρ=0.33, p=0.11).
Drawing upon arguments in human behavioral studies, we reasoned that the first regime reflected the process of 'sensory encoding' [22-26], during which a slower response allows more sensory evidence to be acquired, thereby improving response accuracy. Upon completion of encoding, that is, upon full construction of the (internal) representation of target stimulus [25, 35], additional sampling would not benefit accuracy any further, resulting in the second regime in which accuracy does not trade off against RT. In contrast to the first one, the second regime can involve maintaining the acquired information for later responses, consistent with the visual short-term memory (VSTM, or visual working memory, VWM) idea in human literature (Philip1974; Vogel2006; Smith2009). We, therefore, termed the two regimes, respectively, the sensory encoding stage and the VSTM stage.

To quantify the relationship between conditional accuracy and RT in the sensory encoding regime, we fitted the data with an asymptotic function (conditional accuracy function, CAF; Methods) [18-21]. We then estimated three key metrics of the sensory encoding phase for use in subsequent comparisons between trial conditions (Methods): (1) the peak conditional accuracy (a_{peak}), (2) the timepoint at which conditional accuracy reached its peak (t_{peak}), and (3) the timepoint at which conditional accuracy just exceeded 50% (chance) performance (t_{50}; Methods).

Specifically, we fit the CAF to data from trials of different stimulus sizes (Fig. 2B), and estimated the key metrics of sensory encoding (Methods; all contrasts included). We found that the peak conditional accuracy was significantly modulated by stimulus size (Fig.2C-left; a_{peak}: size 25°, mean ± s.d. =81.3 ± 1.2%; size 35º =88.0 ± 0.7%; size 45º =92.4 ± 0.9%; *, p<0.05, permutation tests with HBMC correction). The time to reach peak accuracy and the time to exceed chance performance were not significantly different across stimulus sizes (t_{peak}, Fig. 2C-middle, size 25°=491 ± 56 ms, size 35°=461 ± 22 ms, size 45°=467 ± 26 ms; t_{50}, Fig. 2C-right, size 25°=190 ± 31 ms, size 35°=221 ± 14 ms, size 45°=193 ± 20 ms)

Next, we examined the effect of target contrast on the key metrics of sensory encoding. We fit the CAF to data from trials of different contrasts (Fig. 2D): contrasts were divided into two levels - low-contrast (contrast levels 1-3, all sizes included), and high-contrast (contrast levels 4-7, all sizes included, Methods). We found that the peak conditional accuracy was significantly modulated by stimulus contrast (Fig.2E-left; a_{peak}: low-contrast = 84.6 ± 0.9%; high-contrast = 89.5 ± 0.6%, p<0.001, permutation test), as was the time to reach peak accuracy (Fig.2E-middle; t_{peak}: low-contrast = 532 ± 30 ms; high-contrast = 412 ± 17 ms, p<0.001, permutation test). We found no significant effect of stimulus contrast on t_{50} (Fig. 2E-right, low-contrast = 213 ± 20 ms; high-contrast = 207 ± 12 ms, p=0.747, permutation test).

Taken together, our findings support the conclusion that stimulus features pose intrinsic limits on discrimination performance: despite the abundance of time available for responding to the target (up to 3s), the best level of accuracy that mice could possibly reach (and the time to reach peak accuracy) were still limited by stimulus features such as size and contrast.

Effect of stimulus duration on mouse discrimination performance and VSTM

In the above experiments, the variation of stimulus features of contrast and size modulated the sensory encoding regime of the conditional accuracy function. Because the stimulus duration in these experiments was fixed at 3s, equal to the time window allowed for mice to respond, there were no trials in which mice responded after stimulus offset. Therefore, it was not possible to test directly whether (and to what extent) mice relied on VSTM to generate responses.

To address this issue and to explore the connection between performance and VSTM, i.e., the second regime of the conditional accuracy function, we next varied stimulus duration systematically, while maintaining the size and contrast fixed (at 25° and 99, respectively). This allowed us to analyze trials on which mice responded after the stimulus disappeared, i.e., trials on which mice may need to rely on the information maintained in VSTM to make their response. Since the information in VSTM is thought to decay over time and exhibit a finite lifetime [36-40], we predicted that for trials in which the mice responded after stimulus offset, the conditional accuracy would decline with longer RTs, allowing an estimate of the duration of VSTM. As an additional benefit, varying stimulus duration allowed us to directly estimate the shortest duration of the stimulus that resulted in above-chance discrimination performance in mice.

We found, first, that stimulus duration significantly modulated discrimination accuracy of mice (Fig.3A, one-way ANOVA, p=0.047), with accuracy decreasing as the stimulus duration decreased (Pearson's ρ=0.74, p=0.01). This effect was driven by a commensurate effect of duration on perceptual discriminability (Fig. S2A; one-way ANOVA, p=0.001; Pearson's ρ=0.74, p=0.01) but not decision criterion (Fig. S2B; one-way ANOVA, p=0.802). There was also a trend of decreased RT as the stimulus duration decreased, although the effect was not statistically significant (Fig.3B, one-way ANOVA, p=0.133; Pearson's ρ=0.86, p<0.001).

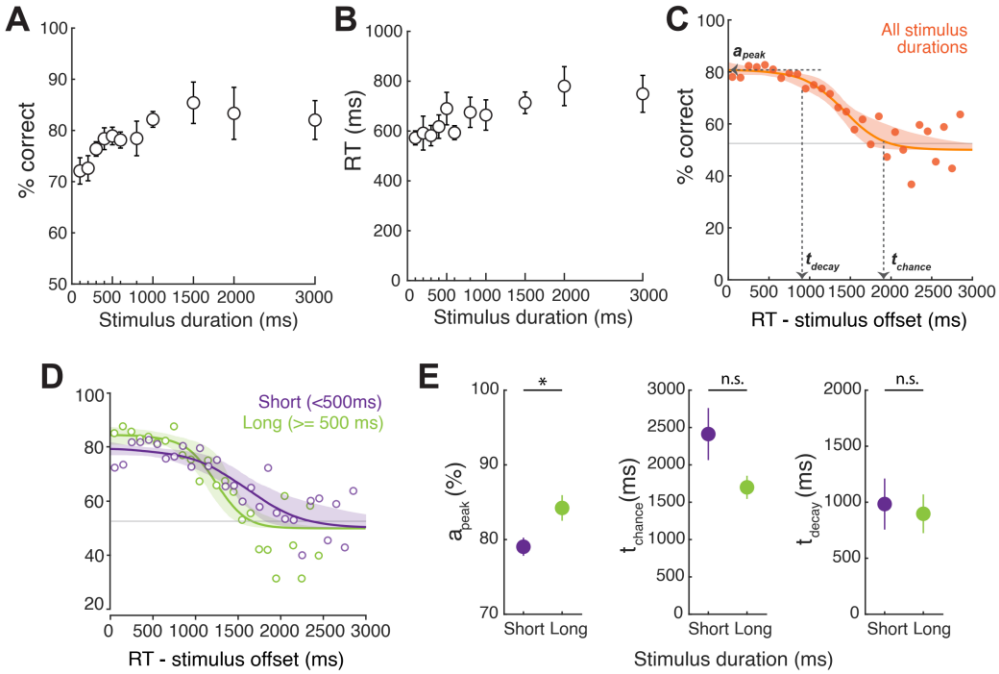

Figure 3. Stimulus duration modulates orientation discrimination performance, and the sensory encoding regime of the CAF. (A) Psychometric plot of discrimination accuracy against stimulus duration. Data: mean ± s.e.m; n= 6 mice. 1-way ANOVA; p=0.047. **(B)** Plot of median reaction time (RT) against stimulus duration. 1-way ANOVA; p=0.133. **(C)** Plot of accuracy as a function of RT bins aligned to stimulus offset (conditional accuracy). Orange dots: Data pooled across all mice and stimulus durations; RT bin size = 100 ms. Orange curve: Conditional accuracy function, CAF (best-fit decaying logistic function; Methods); light orange shading: 95% CI of the fit (Methods). Indicated are one key parameter (a_{peak}) describing the initial, sensory encoding stage, and two key parameters (t_{decay} and t_{chance}) describing the VSTM-dependent stage sensory encoding stage of the CAF - the period during which accuracy and speed of response do not exhibit a tradeoff, and during which accuracy decays with RT (see text; Methods). Peak accuracy (a_{peak}): mean ± s.d. = 80.9 ± 1.2%; time when the performance starts to decay

from a_{peak} (t_{decay}): 931 ± 181 ms; time at which performance decays to chance levels (t_{chance}): 2066 ± 285 ms (Methods). **(D)** CAFs for trials with short (<500 ms, purple) and long (≥500 ms, green) stimulus duration; conventions as in C. **(E)** Plots of key parameters of CAFs in (D) for different stimulus durations. Data show mean ± s.t.d of distribution of bootstrapped estimates (Methods). '*' ('n.s.'): $p<0.05$ ($p>0.05$), paired permutation test (Methods); a_{peak}: p=0.013 (short vs. long stimulus conditions); t_{chance}: p=0.177; t_{decay}: p=0.796. **See also Fig. S2.**

For the conditional accuracy analysis, we focused on the second regime of the CAF, to investigate if mouse discrimination accuracy declined after the stimulus disappeared. We pooled data from all mice (n=6), aligned trials of all stimulus durations by the time of stimulus offset,-and calculated conditional accuracy as a function of RT (after stimulus offset, Fig. 3C). Consistent with our prediction, the conditional accuracy declined with RTs longer than stimulus offset, supporting a decaying VSTM process.

To quantify the time course of decay, we fit the data using a sigmoidal function (Methods), and estimated two key metrics of the VSTM phase for use in subsequent comparisons between trial conditions (Methods): (1) the time point at which the conditional accuracy started to decline (t_{decay}); and (2) the first timepoint at which the discrimination accuracy dropped to a level indistinguishable from the chance (t_{chance}; Methods); indicating that information is no longer available in VSTM.

We found that for the data pooled across stimulus durations, the first instant (t_{decay}) at which the conditional accuracy dropped significantly below the peak accuracy was about 900 ms after stimulus offset (Fig.3C, t_{decay}, mean ± s.d. = 931 ± 181 ms). Conditional accuracy dropped down to chance levels at about 2100 ms after the stimulus offset (Fig.3C, t_{chance} = 2065 ± 285 ms), which allows for estimating the duration of moues VSTM (see last section in Results).

Studies in humans using a variety of techniques have reported a robust effect of stimulus duration on VSTM, called the 'inverse duration effect'. This describes the phenomenon that the longer a stimulus lasts, the shorter is its persistence in VSTM after the stimulus offset [33, 41, 42]. To investigate if the inverse duration effect occurs in mice as well, we split our data into two subsets: (1) trials with stimulus duration < 500ms (100-400ms); and (2) trials with stimulus duration ≥ 500ms (500-2000ms), and repeated the conditional accuracy analysis for the two subsets of trials.

We found that compared to short-stimulus trials, long-stimulus trials tended to have a shorter duration over which the conditional accuracy remains above chance after stimulus offset (Fig. 3DE; t_{chance}: long-stimulus, mean ± s.d. = 1699 ± 155 ms; short-stimulus= 2413 ± 349 ms p=0.177, permutation test). This trend is consistent with the findings of human studies, although the difference does not reach statistical significance in our dataset. There was no difference between two groups in terms of when the conditional accuracy started to decay (t_{decay}, long-stimulus = 897 ± 173 ms; short-stimulus= 984 ± 228 ms p=0.796, permutation test).

Incidentally, the peak accuracy was higher for long-stimulus trials than short-stimulus trials (long-stimulus = 84.2 ± 1.7%; short-stimulus = 79.0 ± 1.2%, p=0.013, permutation test). This could be the result of the sensory encoding stage being terminated prematurely in short-stimulus trials, consistent with the finding from Figure 2A that it takes about 500ms for the stimulus to be fully encoded (i.e., peak conditional accuracy is reached).

Taken together, our results demonstrate that stimulus duration significantly modulates mouse performance in discriminating grating orientation. Discrimination accuracy is lower when stimulus duration is constrained (Fig. 3A), the potential result of a combination of two factors: (1) the *encoding* of visual information is terminated earlier when the stimulus is short (<500ms), consistent with the lower peak conditional accuracy for short vs. long stimuli (Fig. 3E, left), and (2) the duration of *maintenance* of visual information necessary for correct responding (i.e., of VSTM) is limited (Fig. 3C).

Simultaneous presentation of a task-relevant foil ('flanker') modulates conditional accuracy function

The tasks thus far involved the presentation of a single stimulus ('target'), in which mice were challenged by varying stimulus properties. Another factor that can limit animals' performance is the sensory context in which the target is presented. For instance, the co-occurrence of a task relevant foil stimulus with conflicting information from the target stimulus can interfere with perceptual performance, as demonstrated in the classic flanker task in humans [43, 44]. In this task, the target is always presented at a fixed location, but competing stimuli with congruent or incongruent task-relevant information are presented at flanking locations. In a recent study in mice, using a mouse version of the flanker task, we demonstrated similar result: the presence of a foil stimulus with conflicting information ('incongruent flanker') significantly impaired mouse discrimination performance, but not the presence of a foil stimulus with congruent information ('congruent flanker'; Fig. 4AB; Methods). The stimuli in this task were presented for 1000 ms. Here, we re-analyzed that dataset with the conditional accuracy analysis to investigate whether the performance reduction observed was due to the interference by the incongruent flanker in the process of sensory encoding, or in the maintenance of target information in VSTM.

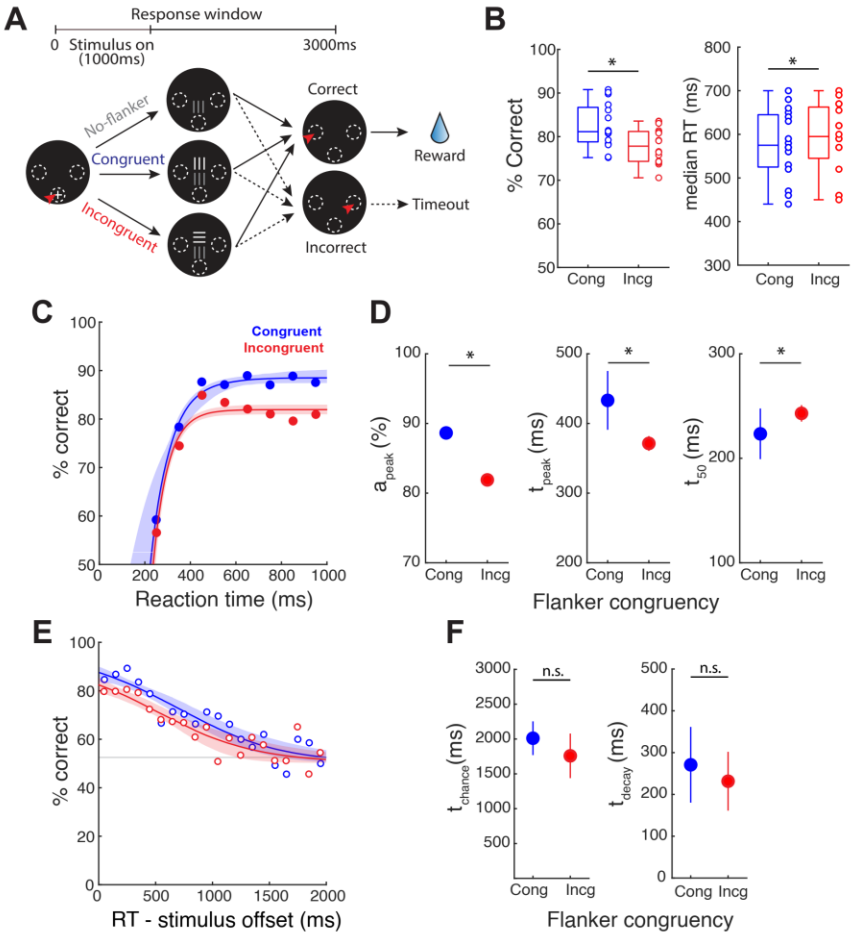

Figure 4. Incongruent flanker reduces target discrimination accuracy and affects the sensory encoding regime of the CAF. (A) Schematic of the flanker task; target grating is always presented at the lower location; a second 'flanker' grating (orthogonal orientation – incongruent flanker, or same orientation – congruent flanker) is presented simultaneously, and always at the upper location; contrast of flanker is systematically varied (adapted

from [16]). All other conventions as in Fig. 1. Plots represent results from new analyses applied to previously reported data [16] after collapsing across all flanker contrasts (Methods). **(B)** Left panel: Comparison of performance between trials with incongruent vs. congruent flanker. $p<0.001$, signed rank test. Right panel: Comparison of median RT between trials with incongruent vs. congruent flanker. $p=0.019$, signed rank test. **(C)** CAFs of the sensory encoding stage; data correspond to trials with RT < stimulus offset, i.e., 1000 ms. Blue: trials with congruent flanker. Red: trials with incongruent flanker. **(D)** Plots of key parameters of CAFs (sensory encoding stage) for trials with congruent vs. incongruent flanker; a_{peak} (left), t_{peak} (middle), and t_{50} (right). Data show mean ± s.t.d of distribution of bootstrapped estimates. '*' ('n.s.'): $p<0.05$ ($p>0.05$), permutation tests followed by HBMC correction, congruent vs. incongruent flanker conditions (Methods). a_{peak}: $p<0.001$; t_{peak}: $p=0.01$; t_{50}: $p=0.022$. **(E)** CAFs of the VSTM-dependent stage; data correspond to trials with RT > stimulus offset (1000 ms), aligned to stimulus offset. Blue: trials with congruent flanker. Red: trials with incongruent flanker. **(F)** Plots of key parameters of CAFs (VSTM-dependent stage) for trials with congruent vs. incongruent flanker; t_{chance} (left) and t_{decay} (right). Conventions and statistical methods as in D. t_{chance}: $p=0.505$; t_{decay}: $p=0.410$.

We pooled trials from all mice into two groups based on their flanker congruency – i.e., congruent vs. incongruent. Following that, for each group, we sorted the trials based on their RT. Trials with RT shorter than the duration of stimulus (1000ms), i.e., trials in which mice responded before the stimulus ended, were used to investigate the sensory encoding regime (per the approach used in Figure 2). Separately, trials with RT longer than the duration of stimulus, i.e., trials in which mice responded after stimulus offset, were used to investigate the VSTM stage (per the approach used in Figure 3).

We found that in the sensory encoding regime (Fig. 4CD), the peak conditional accuracy for incongruent trials was significantly lower than that of congruent trials (Fig. 4D-left; congruent = 88.6 ± 0.8%, incongruent = 81.9 ± 0.5%; $p<0.001$, permutation test), and the time at which performance just exceeded the 50% (chance) level was longer for incongruent trials (Fig. 4D-right; t_{50}: congruent = 223 ± 24 ms; incongruent = 243 ± 8 ms $p=0.022$, permutation test). The time to reach peak accuracy was, however, shorter for incongruent trials (Fig. 4D-middle; t_{peak}:congruent = 433 ± 42 ms; incongruent = 371 ± 11 ms; $p=0.01$, permutation test), consistent with the higher a_{peak} (Fig. 4D-left) combined with similar slopes of the CAF (Fig. 4C).

By contrast, there was no effect of flanker congruency on the time course of decay of conditional accuracy following offset of the target and flanker stimuli (i.e., the VSTM stage). The time at which conditional accuracy dropped to chance was not different between congruent and incongruent flanker trials (Fig. 4EF; t_{chance}: congruent = 2011 ± 242 ms; incongruent = 1759 ± 320 ms, $p=0.505$, permutation test), nor on the time at which conditional accuracy dropped just below a_{peak} (t_{decay}: congruent = 271 ± 91 ms; incongruent = 232 ± 70 ms, $p=0.410$, permutation test).

In sum, we found that the interference in performance due to the incongruent flanker mainly impacted the process of sensory encoding (a_{peak}; as if weakening the target), but not the VSTM stage.

Relative target contrast (target:foil) modulates mice's performance in visual target selection

We next challenged mice with a visual search task, which involved added complexity compared to the flanker task. Here, after a trial was initiated, the target grating could be presented either alone ('singleton trial') or together with a second grating (foil; 'search trial'; Fig. 5A). However, unlike the flanker task, (a) the target was defined as the stimulus of higher contrast (as opposed to the stimulus at a particular location), (b) the location of the target was randomized on a trial-by-trial basis, and (c) the orientation of the foil was always orthogonal to the orientation of the target (chosen randomly on each trial to be either horizontal or vertical). The relative contrast of the target:foil was varied systematically from 1.73 to 64.8 following a contrast morphing protocol: the contrast of one stimulus decreased while that of the other increased over the range of contrasts. In the singleton trials, the contrast of the target was varied over the same range (Fig. 5A-right; Methods). This task adds complexity over the flanker task as it not only contains a task-relevant competitor, but in addition, involves uncertainty of the target location.

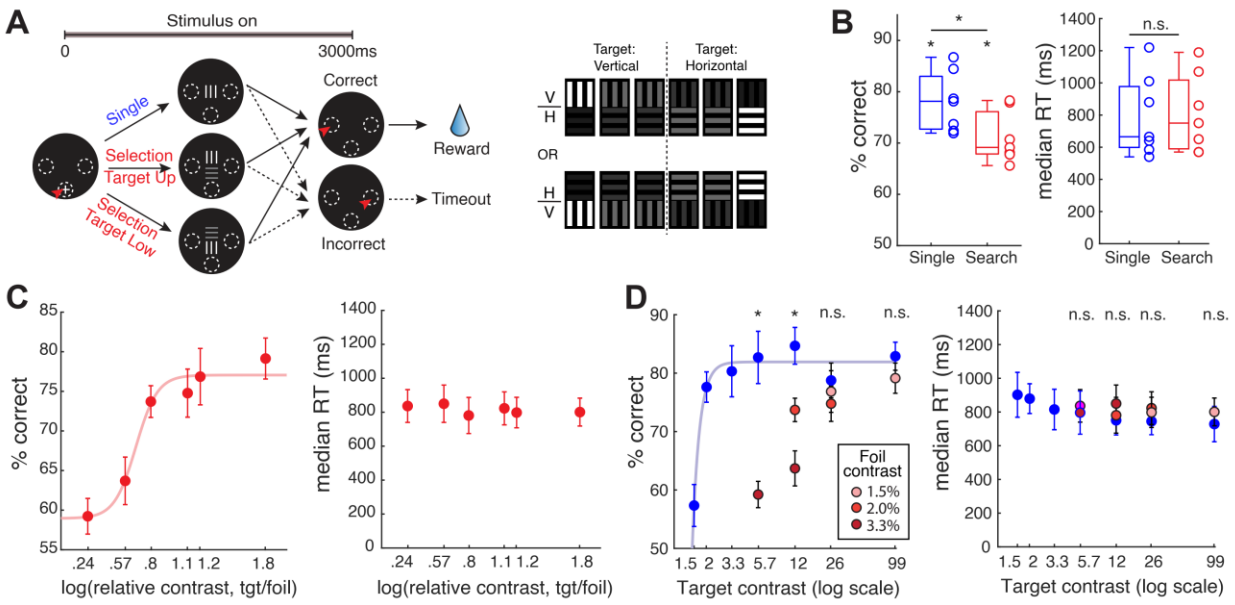

Figure 5. Foil contrast modulates target discrimination accuracy in visual search task. (A) Left: Schematic of rudimentary visual search task. Two gratings of different contrasts and orthogonal orientations are presented simultaneously; the target is defined as the grating of higher contrast. Relative contrast of target and foil are varied following a contrast morphing protocol. Right: Schematic of contrast morphing protocol: as the contrast of vertical grating decreases (from left to right), the contrast of paired horizontal grating increases. Therefore, the target can be either the vertical (left half of example pairs) or the horizontal grating (right half of example pairs). Additionally, the target (higher contrast) grating can occur either at the upper location or at the lower location with equal probability. Only a subset of grating pairs are shown here as examples; all other conventions as in Fig. 1. **(B)** Effect of foil on response accuracy (left) and median RT (right); data from trials pooled across all relative contrast conditions. Single-stimulus trials vs. search trials; response accuracy: p=0.016; signed rank test ('*'); median RT: p=0.484, signed rank test ('ns'). Response accuracy was additionally significantly higher than chance (50%) for search trials (p=0.016, signed rank test). **(C)** Psychometric curve of response accuracy (left) and median RT (right) plotted against relative contrast of target:foil. Red curve: best sigmoidal fit (Methods). Left: p<0.001, 1-way ANOVA. Right: p=0.996, 1-way ANOVA. **(D)** Comparison of discrimination performance when target was presented alone (blue data points) vs. when it was presented with foil (red data points). Left: Accuracy. Right: median RT. Darker shades of red: higher contrasts of foil. '*' ('n.s.'): p<0.05 (p>0.05), Kruskal-Wallis tests followed by HBMC correction (Methods). **See also Fig. S3.**

We found that mice were able to learn this search task well. Their discrimination accuracy was significantly higher than chance in both trial types (Fig. 5B, left panel; singleton trials, median=78.1%, [72.0, 84.3], p=0.016, signed rank test; search trials, median=69.1%, [64.2, 74.0], p=0.016, signed rank test). Compared to the singleton trials, mouse discrimination accuracy was significantly lower in the selection trials (p=0.016, signed rank test), indicating impairment in performance due to the presence of the foil.

The search task also yielded a classic psychometric curve of performance (Fig. 5C). Mouse discrimination accuracy was dependent on the relative target contrast (Fig. 5C-left, one-way ANOVA, p<0.001): as the relative target contrast decreased, discrimination accuracy decreased as well. This reduction was not simply due to that the target being dimmer in those pairs with low relative target contrast: Comparing the discrimination accuracy of the same target when it was presented alone, versus when it was presented with a foil, demonstrated that the deterioration in accuracy was due to and dependent on the contrast of

foil (Fig. 5D-left, Kruskal-Wallis test with HB correction). Additionally, these effects on discrimination accuracy were driven by commensurate changes in perceptual discriminability (Fig. S3AB, left panel) rather than changes in decision criterion (Fig. S3AB, right panel). In contrast to discrimination accuracy, RT was minimally affected by the presence of foil (Fig. 5B, right panel: p=0.48, signed rank test) nor by the relative target contrast (Fig. 5C - right: one-way ANOVA, p=0.996; Fig, 5D - right).

We next performed the conditional accuracy analysis on this dataset and analyzed the sensory encoding stage following the approach outlined in Figure 2 (Fig. S3DE; the stimulus duration was equal to the response window, 3s, in this task). We found that search trials had a lower peak accuracy than singleton trials (Fig. S3E-left; a_{peak}: singleton trials = 80.0 ± 1.9%, search trials = 73.1 ± 1.4%, p=0.004, permutation test). The time to reach peak accuracy was not statistically different between the two conditions (Fig. S3E-middle; t_{peak}: singleton trials = 417 ± 44 ms, search trials = 371 ± 42 ms, p=0.418, permutation test), and neither was the time at which performance just exceeded the 50% (chance) level (Fig. S3E-right; t_{50}: singleton trials = 257 ± 24 ms, search trials = 274 ± 13 ms, p=0.434, permutation test).

In addition, search trials with low relative contrast of the target (trials corresponding to first three relative contrast values) had lower peak accuracy than those with high relative contrast (trials corresponding to last three relative contrast values; Fig. S3F, S3G-left; a_{peak}: high relative contrast search trials = 80.8 ± 2.0%; low relative contrast search trials = 65.8 ± 2.2%; p<0.001, permutation test). There was no significant effect of relative contrast on the time to reach peak accuracy (Fig. S3G-middle; t_{peak}: high relative contrast = 415 ± 91 ms; low relative contrast = 342 ± 66 ms; p=0.349, permutation test), nor on the time at which performance just exceeded the 50% (chance) level (Fig. S3G-right; t_{50}:high relative contrast = 270 ± 18 ms; low relative contrast = 269 ± 32 ms; p=0.944, permutation test)

In order to make the search task more challenging, and to potentially investigate the effect of the foil on the VSTM-dependent regime, we reduced the stimulus duration to 800 ms (from 3000 ms). However, of the 7 mice that we attempted to train on this more difficult search task, we found that only 2 reached the criterion for successful demonstration of learning in this task, namely, accuracy >= 70% across all trial types (Fig. S3H- filled dots; S3I; Methods). This was in contrast with the original search task (Fig. 5B) in which all mice consistently exhibited >70% accuracy in single target trials.

Thus, mice were able to learn the original search task (with stimulus duration = 3s), and their performance was systematically affected by the presence and strength (contrast) of the foil, which significantly modulated the sensory encoding stage of the CAF. However, when stimulus duration was reduced (800 ms), the task appeared to exceed the capability of most mice (5/7).

Stimulus onset delay modulates mice's performance and reveals impulsivity of mice

Finally, we were interested in another 'limit' related to mouse performance. We wished to assess quantitatively the extent to which mice are naturally able to withhold responding when there is a delay in the arrival of information pertinent to the task, i.e., their ability to adjust the timing of their responses adaptively to the temporal statistics of the stimulus. To this end, we trained mice on the single stimulus discrimination task (as in Fig. 1), and systematically varied the delay between trial initiation and onset of the target (stimulus onset delay).

We found that discrimination accuracy was not modulated by the stimulus onset delay (Fig. 6A; one-way ANOVA, p=0.337) although there was a trend towards lower performance for longer delays. Similarly, there was no effect of stimulus onset delay on perceptual discriminability or decision criterion (Fig. S4AB). We next examined the effect of stimulus onset delay on RT. If animals are able to withhold responding until the target comes on and information about the orientation becomes available, then we

would expect the RT distributions, and therefore the median RT to remain unchanged as a function of delay (because RT is measured with respect to target onset). However, we found a significant *reduction* in the median RT as the stimulus onset delay increased (Fig. 6B; one-way ANOVA, $p<0.018$; Pearson's ρ = -0.99, p=0.023) – mice were responding earlier when the stimulus onset delay was longer. To gain insight into this puzzling finding, we replotted the response time data, but now, calculating reaction time (RT_{INI}) as the time of response from trial initiation (instead of from stimulus onset; Fig. 6C). Our motivation for this analysis was the hypothesis that perhaps mice are insensitive to stimulus onset delay. If so, we would expect the distributions of RT_{INI} (and the median RT_{INI}s) to be nearly identical across delays, thereby explaining the decrease in RT as a function of delay. We found that RT_{INI} showed an increasing trend with stimulus onset delay (Fig. 6D; Pearson's $\rho = 0.999$, p=0.035), but with the magnitude of the change in median RT_{INI} being less than the magnitude of change in delay (average Δ median RT_{INI} = 36 ms vs. Δ delay = 100 ms between the first two delays, and average Δ median RT_{INI} = 43 ms vs. Δ delay = 100 ms between between the second two delays).

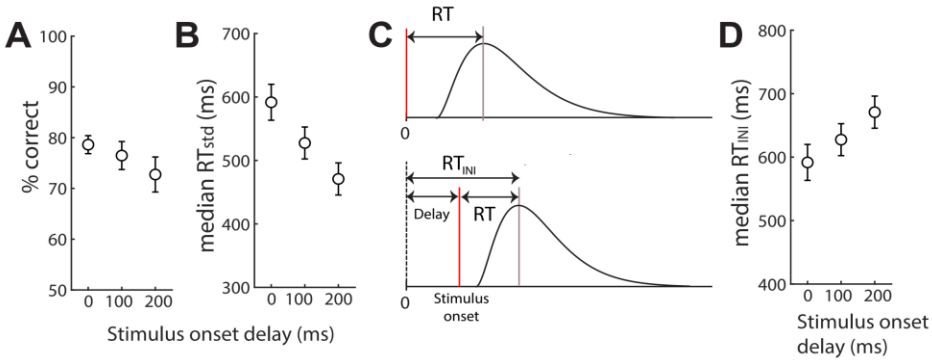

Figure 6. Stimulus onset delay modulates RT of orientation discrimination and allows quantification of impulsivity in freely behaving mice. (A) Plot of response accuracy against stimulus delay; p=0.337, 1-way ANOVA; ρ= 0.988, p=0.098. **(B)** Plot of median RT against stimulus delay; RT calculated from stimulus onset; p=0.018, 1-way ANOVA; ρ= 0.99, p=0.023. **(C)** Schematic illustrating distributions of response times under two conditions: Top panel – no delay; Bottom panel: non-zero stimulus onset delay. 0 indicates train initiation, red line indicates stimulus onset. Bottom panel illustrates two different ways in which response times are calculated here: with respect to stimulus onset (RT; the standard method), or with respect to trial initiation (RT_{INI}). **(D)** Plot of median RT_{INI} against stimulus delay. RT_{INI} is slower for greater onset delay (Pearson's $\rho = 0.999$, p=0.035), indicating that mice are able to sense the delay and withhold their response. However, they do not withhold for the entire duration of the delay. The ratio of duration for which they withhold over delay duration is defined the impulsivity index (=0.6 for mice). **See also Fig. S4.**

Thus, mice appear to be able to sense stimulus onset delays and withhold their responses (resulting in longer RT_{INI}s), but are unable to do so for the full duration required (resulting in shorter RTs). As a result, their responses were often premature (or 'impulsive'), with movement initiated before the stimulus was even presented, causing them to guess more often.

These data allowed us to estimate impulsivity quantitatively with an 'impulsivity index': ImpI = 1 – (duration that mice waited /duration of optimal wait time). ImpI = 0 would indicate that animals are non-impulsive (optimal), higher positive values of ImpI would indicate that animals are more impulsive, with ImpI=1 indicating that animals that are unable to withhold responding at all (maximally impulsive). In the case of our mice, ImpI was ~0.6.

In line with the analysis approach in previous tasks, we examined the effect of stimulus onset delay on the conditional accuracy function to characterize both the sensory encoding stage as well as the VSTM stage in this task. We found that there was no effect of stimulus onset delay on the encoding regime (Fig. S4DE; a_{peak}: no delay = 83.4 ± 1.9%; delay = 83.1 ± 2.6%; p=0.921, permutation test; t_{peak}: no delay =

430 ± 64 ms; delay = 417 ± 136 ms; p=0.887, permutation test; t_{50}: no delay = 193 ± 47 ms; delay = 147 ± 47 ms; p=0.105, permutation test), nor on the VSTM regime (Fig. S4FG; t_{chance}: no delay = 1611 ± 337 ms; delay = 2106 ± 599 ms; p=0.064, permutation test; t_{decay}: no delay = 898 ± 217 ms; delay = 641 ± 266 ms; p=0.156, permutation test).

Thus, varying the stimulus onset delay did not affect either the sensory encoding or the VSTM regimes of response, but revealed the impulsivity of mice and allowed a quantitative characterization of impulsivity in mice.

Estimates of motor response time, visual stimulus sampling period and length of visual short term memory

The conditional accuracy analysis yielded estimates of key time points within the two regimes of the CAF, namely, t_{peak} and t_{50} for the sensory encoding regime, and t_{decay} and t_{chance} for the VSTM-dependent regime. These values, which were all measured as reaction times, included two fixed overheads: (a) the sensory processing delay -- time taken for the visual periphery to transduce and relay sensory information to visual brain areas, i.e., neural response latency), as well as (b) the motor execution delay -- the time between the brain making the decision and the animal reporting its choice, corresponding to the actual time for movement of the mouse's head and body to achieve a nose-touch in this case. To obtain accurate estimates of the durations of the underlying decision processes, it would be important to subtract away the fixed sensory and motor "overheads".

An analysis method that permits the combined estimation of this fixed overhead time is drift diffusion modeling (DDM [45, 46]; Methods). The DDM model is fit to the full RT distributions obtained in 2-choice behavioral tasks to estimate four different parameters corresponding to potential psychological variables underlying performance. One of the parameters is exactly the quantity we are interested in, namely, the fixed 'overhead' time accounting for both sensory and motor execution delays, and is typically termed the 'non-decisional constant' in DDMs. The other three parameters are: the rate of evidence accumulation (drift rate), the distance between the subject's decision boundaries, and the internal bias of the subject towards one of the choices.

We fit the DDM to the left-light choice RT distributions of mice in several of the tasks in this study, and in each case estimated the values of the four parameters [47, 48]. Specifically, we quantified the non-decisional constant in each case (Fig. 7). We found that in each case, the non-decisional constant was not modulated significantly by the variable of interest in that task (Fig. 7A, size and contrast of target: 2-way ANOVA, size: p=0.308, contrast: p=0.523; interaction: p=0.931; 7B, flanker congruency and contrast: 2-way ANOVA, congruency: p=0.343, contrast: p=0.998; interaction: p=0.993; Fig. 7C, relative contrast of target:foil: 1-way ANOVA, p=0.269). Additionally, the mean values of the non-decisional constant were nearly identical across tasks (7A- 210 ms ± 9.6 ms, 7B- 212 ms ± 6.9 ms, 7C: 188 ms ± 9.5 ms). Based on these results, we estimated that the fixed overhead for mice across all our tasks was approximately 200 ms.

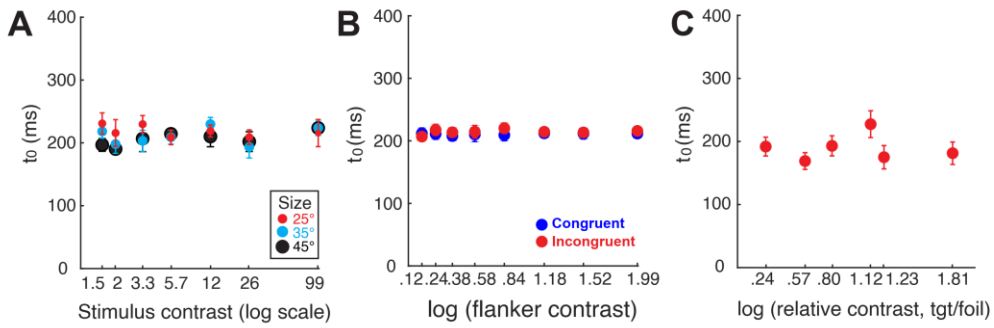

Figure 7. Drift diffusion modeling reveals nearly fixed time for combined sensory latency and motor execution (non-decisional time, t_0) across tasks. (A) Estimates of t_0 by drift diffusion modeling as a function of stimulus size and contrast (same dataset as in Fig. 1; Methods). 2-way ANOVA, p=0.523 (contrast), p=0.308 (size), p=0.931 (interaction). **(B)** Estimates of t_0 as a function of flanker congruency and flanker contrast in flanker task (same dataset as in Fig. 4; Methods). 2-way ANOVA, p=0.343 (flanker congruency), p=0.998 (flanker contrast), p=0.993 (interaction). **(C)** Estimates t_0 as a function of relative contrast of target:foil in visual search task (same dataset as in Fig. 5; Methods); p=0.269, 1-way ANOVA. **See also Fig. S5.**

With this information, we then estimated the window for sensory encoding (temporal integration) as t_{peak} - 200 ms = 300 ms (Fig. 2A; t_{peak} ~= 500 ms). We also estimated the duration of visual short term memory as the duration of the period starting from 200 ms after stimulus offset (the last instant at which a response could have been initiated with the stimulus still on the screen), to 200 ms before t_{chance} (the last instant at which responses that are better than chance was initiated). Thus, we estimated the duration of mouse VSTM as t_{chance} – 2*200ms = 1700 ms (Fig. 3C; t_{chance} ~= 2100 ms).

DISCUSSION

In this study, we explored the limits of the visual discrimination performance in freely behaving mice by systematically manipulating stimulus size, contrast, duration, delay and sensory context in a 2AFC, orientation discrimination task. The resulting psychometric curves revealed that mouse discrimination performance was robust to key stimulus features, and persisted in the presence of a foil stimulus. By revealing parallels in perceptual processes in humans and mice, our results indicated that although mice have poorer visual acuity compared to primates [1, 11], their visual perceptual abilities may be underrated. These findings establish a quantitative, psychophysical foundation for the future study of the neural basis of visually guided behavior in mice.

Performance at small stimulus feature values

In general, shrinking the size of stimulus, lowering the stimulus contrast, or shortening the stimulus duration all caused deterioration of discrimination performance (as expected). However, mice were able to respond to visual stimuli that were smaller, more brief, and in general, more demanding than those typically used in studies of vision and visually guided behavior in mice [6, 9, 11, 12, 49]. Based on the ranges of values that we tested, mice were able to discriminate vertical versus horizontal gratings at a stimulus size as small as 25º, and for that small size, they could discriminate a stimulus as short as 100 ms (at full contrast), or with as low contrast as 1.5 (at duration of 3s). Even at small stimulus sizes, low contrasts, and short durations, mice were able to perform consistently better than chance, with an accuracy of 70-75%.

Performance at large stimulus feature values

At the other end of the range, performance plateaued at 93% for a size of 45°, suggesting that the full field stimuli that have been previously used in mouse visual studies may be effectively replaced by 45° large stimuli without appreciable loss in performance. Similarly, performance plateaued at 83% for stimulus duration ≥ 1000 ms, indicating that stimuli longer than 1000 ms may not be needed to test mouse behavior effectively in single-stimulus discrimination tasks. Finally, with respect to stimulus contrast, although mouse performance generally improved as the contrast increased, there appeared to be a dip in performance as the stimulus reached full contrast. Such inverted U-shaped performance curves as a function of contrast have been reported previously in mice in a go/no-go task [9] . A potential explanation that has been offered is that this dip is due to the variability of stimulus contrast inherent when multiple values are tested [9]. Since the visual system is known to adapt to the range of stimulus contrast for best encoding [50], it is possible that the large variability in contrast values in an experiment that parameterizes contrast makes the full-contrast stimulus unfavorable because of signal saturation. Our data are consistent with this idea: mouse performance to full-contrast gratings was worse when the stimuli were intermixed with other contrasts (Fig.1C, mean accuracy= [79%, 81%, 85%] corresponding the three stimulus sizes), compared to when they were presented at a single contrast (Fig. S1F, mean accuracy= [85%, 90%, 94%]). For all the stimulus parameters tested, changes in discrimination accuracy were accompanied by changes in sensitivity rather than decision criterion, indicating that the manipulations all modulated aspects of the perceptual process.

Orientation discrimination: rapid, discrete perceptual processing or fast integration-to-threshold?

Measurements of RT in these tasks revealed an intriguing result: median RT was approximately 600 ms in nearly all experiments, and, notably, was either unaffected by stimulus manipulations, or when it was, changed within a range of approximately 100 ms. This result is distinct from findings in perceptual decision-making tasks such as the random dot motion discrimination (RDM) task (in primates), in which increasing task difficulty substantially slowed down RT, by more than 500 ms [51, 52]. Such tasks are well accounted for by an accumulation-to-threshold process underlying perceptual decision-making: Waiting longer to respond in tasks that are more perceptually difficult (i.e., with slower rates of sensory evidence) would result in accumulation of more evidence, and thereby benefit performance. However, our result of largely flat RTs is similar to findings from odor discrimination tasks in rodents [53, 54], in which increasing task difficulty did not significantly increase RT. These tasks, in which rats exhibited nearly constant odor sampling times of ~300 ms, have been shown to not be accounted for by accumulation-to-threshold processes, even with short integration timescales and changing thresholds [55]. Rather they have been shown to represent a distinct, rapid form of perceptual decision-making that involves discrete sampling of low-level sensory information [55].

Our drift diffusion modeling helped disambiguate whether the performance of mice in our tasks was better explained as an integration-to-threshold process (as in monkey RDM tasks) or a rapid form of perceptual process with snapshot sampling (as in rad odor discrimination tasks). We found that across our tasks, drift rates systematically changed with task difficulty (Fig. S5AEI). Notably, decreases in drift rate (with increasing task difficulty) were associated with a lowering of the decision threshold (Fig. S5DHL), thereby together accounting for the finding of largely constant median RTs. In other words, although the behavioral phenotype of nearly constant RTs across task difficulty levels is seemingly similar to snapshot sampling in rodent odor discrimination, the visual discrimination tasks studied here are more closely aligned with classic integration-to-threshold perceptual decision-making tasks, but with integration just occurring on much shorter timescale (temporal integration window = 300 ms)[56, 57].

Conditional accuracy analysis

The conditional accuracy function (CAF; [18-21]) links the two commonly used metrics of behavioral performance, namely, accuracy and reaction time. It represents the accuracy corresponding to each bin of the RT distribution (RTD). As a result, application of the conditional analysis can help decompose observed change in response accuracy following any experimental manipulation, into changes in the

CAF, in RTD, or both. For instance, we found that although manipulating stimulus size/contrast, stimulus duration, the visual context of target stimulus (through presentation of a foil), and pre-stimulus delay all cause changes in response accuracy, they do so via different routes. Manipulating stimulus size/contrast and presenting a foil mainly impact the sensory encoding regime of the CAF (via change in peak accuracy). Varying the stimulus duration involves changes in both the CAF and RTD. Adding a pre-stimulus delay, on the other hand, primarily changes the RTD. In addition, the specific pattern of change in CAF (i.e., a_{peak}, and t_{chance}) provided insights into various cognitive variables that might be affected by the experimental manipulation. This is discussed in the following selections.

Speed-accuracy tradeoff

The conditional accuracy analysis yielded insights into the timescale of visual integration of mice. Across tasks, we found that performance improved with RT for RTs under 500 ms. In other words, for mice, after 300 ms of exposure to the visual stimulus (i.e., RT=500 ms; see Results and Fig. 7), sacrificing speed by sampling longer or deliberating more, did not yield additional perceptual benefits. Together with the estimate from drift diffusion modeling of a non-decisional time constant of ~200 ms, this revealed that the timescale of sensory/evidence integration was ~300 ms for mice in these visual tasks.

The conditional accuracy analysis also offered another window into the result of consistent RT (minimal RT change) across tasks. Since task manipulations did not change the time to peak accuracy (t_{peak}), an optimal strategy would be to respond with RTs centered around t_{peak}: responding earlier than t_{peak} sacrifices accuracy, while responding later than t_{peak} wastes time (reducing the number of trials and amount of potential reward within a session). Consistent with this expectation, the median RT was between 500 ms– 600 ms across tasks, suggesting that mice are operating close to the optimal point in the speed-accuracy tradeoff.

VSTM: Mice vs. humans

Different labels have been used for the time-dependent, labile internal representations of stimuli, such as iconic memory [32, 58], sensory storage [27, 59-61], perceptual memory [34], VSTM [30, 31], visual working memory (VWM, [25, 35, 38, 62]), visible persistence [33] and so on. The duration of VSTM in humans has been estimated to range from 250ms [63] to 1s [27] or even longer [29, 30], and may be depending on the demands placed on the subjects [32] and on the stimulus feature being tested [62]. Here, the combination of conditional accuracy analysis and drift diffusion modeling yielded quantitative insights into the VSTM of mice. We found that mouse conditional accuracy (in the context of grating orientation) gradually decayed after stimulus offset, with the decay starting at about 900ms after stimulus offset. The decay lasted till about 2100ms after stimulus offset, which yielded an estimate of 1700 ms (after subtracting fixed sensory + motor response delays), as the duration of the decaying trace of the stimulus or VSTM, paralleling the estimate from human studies [27, 29].

The duration of VSTM is not its sole limiting feature. The process that encodes information into the form of a VSTM trace is also limited in its capacity [64-70]. It has been shown that the VSTM trace is formed in humans within the first 200-300ms of stimulus presentation [22, 24-26]: masking the stimulus after that time does not impair human subjects' performance. In the previous section, we estimated that mice needed ~300 ms to fully encode a grating stimulus, similar in magnitude to the estimate from human studies.

An inverse relationship has been reported between the duration of stimulus and the duration of information persistence in VSTM after stimulus offset – shorter persistence for longer stimulus durations. We found that mice exhibited a trend for a similar effect in the single target orientation discrimination task. Thus, consistent with another recent report [71], our results establish quantitative parallels between many aspects of human and mouse visual perceptual performance.

The limited duration of VSTM also provided a plausible explanation for the observation that mice tended to respond faster when the stimulus was shorter (Fig.3B), even when the task did not require them to do so (the allowed response window was 3s irrespective of stimulus duration). Since the target information retained in VSTM starts to fade away after the stimulus offset, an optimal adaptation to briefly presented stimuli would be to respond sooner, thereby driving RT to be shorter. Consequently, the 'time pressure' to respond quickly before the VSTM trace of the target fades away is likely to drive the overall (median) RT to smaller values.

Flanker and visual search tasks

We examined mouse discrimination performance in tasks in which the target was presented together with a foil: in one case, the target always occurred at a fixed location, and in the other, target and foil positions were chosen randomly on any given trial. The former is a version of the classic flanker task in humans [43, 44], and the latter, a rudimentary version of the classic search task in humans [72, 73]. (Tasks of both kinds have been used to study selective spatial attention in humans, and we have recently shown the efficacy of the former for studying exogenous capture of spatial selective attention in mice [16]). In the search task, the target is identified both by its salience (higher contrast) and by its behavioral relevance (the learned rule that the higher contrast stimulus yields reward [74]). Thus, both exogenous as well as endogenous influences drive target 'search' and behavioral performance. Consistent with this interpretation, we find that the psychometric curve of performance cannot be explained simply by the varying target contrast: Direct comparison of the performance of paired target-foil presentation with single target presentation (of matched contrast) revealed a foil contrast-dependent performance impairment, consistent with it serving as a progressively more powerful distracter for the animal's spatial attention [75-77].

Stimulus onset delay and impulsivity

Adding a delay before stimulus onset impaired mice's performance. Specifically, the delayed onset of stimulus induced a greater proportion of premature responses (movement initiated before the stimulus was presented), with mice guessing more often. Nonetheless, our results revealed that mice were able to sense the delay and withhold their responses for a short period of time, but just not long enough to fully offset the delay. First, these results suggest that such inherent impulsivity may be countered against by further training, a conclusion supported from work in head-fixed mice in which they were trained to wait during a delay period [15], or withhold licks during a delay period [9]. Second, these results allowed the definition of a quantitative metric for impulsivity, one that depended on the animals withholding responses until information that determines which response to produce is available, rather than depending on withholding responses (after all the information is available) until a 'go' cue is presented, or on the ability to stop a response that is underway [13]. We called this metric the impulsivity index, and estimated it to be 0.6 for mice. It can be used readily to estimate and compare impulsivity across other animals.

METHODS

Animals. All mice were of the C57Bl6/J strain, and were purchased from the Jackson Laboratory. Upon arrival, mice were housed in a colony where temperature (~75F) and humidity (~55%) were controlled on a 12:12h light:dark cycle. Animals were allowed to acclimate for at least one week, with *ad libitum* access to food and water before water regulation was initiated. Experiments were all carried out in the light phase. All procedures followed the NIH guidelines and were approved by the Johns Hopkins University Animal Care and Use Committee (ACUC).

Water regulation. Mice were water-restricted following protocols described by Guo, Hires [78] with a few modifications described previously [16]. Briefly, mice were individually housed, 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.

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). Two custom modifications were introduced that limited the area of the touchscreen available for exploration by the freely behaving mice, thereby minimizing false-alarm triggers due to accidental touches. First, mice were placed within a clear plexiglass tube that ran from the touchscreen to the reward port. The diameter of the tube (5 cm) was large enough to allow mice to run back and forth from the touchscreen to the reward port, to groom and to behave naturally. Second, a thin plexiglass mask (3 mm thickness) was placed 3 mm in front of the touchscreen with three apertures corresponding to the locations at which a mouse was allowed interact with the screen by a nose-touch (Fig. 1A). The apertures, each 1 cm in diameter, were drilled in the mask in an inverted triangle configuration: '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 (Fig. 1A). All experimental procedures were executed using control software (K-limbic, Med-Associates).

Visual stimuli. Visual stimuli (bright objects on a dark background; background luminance = 1.32 cd/m^2) were generated using MATLAB (Mathworks) and imported into the K-Limbic system as jpeg images. 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. The experimental stimuli were oriented gratings (horizontal or vertical), generated using a square wave (24 pixels/cycle; within the range of spatial frequencies shown to be effective for mice [10]). The dark phase of the cycle was black (luminance, L_{dark}= 1.32 cd/m^2; same as the background), and the bright phase was varied between 1.73 cd/m^2 and 130 cd/m^2 depending on the tasks (see below). The contrast of each grating stimulus was calculated as the ratio of luminance of its bright phase (L_{bright}) over its minimal luminance (i.e., dark phase); contrast = L_{bright}/L_{dark}. The size of the stimulus was also varied depending on the task, ranging from 60 pixels x 60 pixels to 156 pixels x 156 pixels, which subtended 25-65 visual degrees at a viewing distance of 2 cm from the screen.

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 behavioral session began with a 10 sec acclimation period, during which mice were allowed to explore the environment with the lights on and to retrieve a bolus (10 µL) of 'free' water at the reward port. Following this, lights were shut off and the zeroing cross to start the first trial appeared on the screen. The cross flashed once every 10 sec until touched, and the flash was accompanied by a short beep of 600 Hz for 30 ms, to induce the mouse to approach and begin the trial. Upon trial initiation, the cross vanished, and the visual stimulus (or stimuli) were immediately presented (typically, but see stimulus onset delay task below), for a duration of 0.1-3s depending on the task (see below).

Mice were trained to report the information contained in the target grating, namely, its orientation, by nose-touching within the correct response aperture (vertical target grating → nose-touch in left response aperture; horizontal target grating → nose-touch in right response aperture). A correct response triggered a tone (600 Hz, 1 sec), the turning on of the magazine light above the reward port, and the delivery of 10 microliters of water at the reward port. Mice turned away from the screen, ran to the liquid well, consumed the reward, and ran back to face the touchscreen in order to begin the next trial. Mouse head entry into the reward port was detected by an infrared sensor which caused the magazine light to turn off, and the zeroing cross (for the next trial) to be presented on the touchscreen. An incorrect response triggered the turning on of both the house light and the magazine light for 5-s as a timeout; the next trial could not be initiated until the end of timeout. A failure to respond within 3s of stimulus presentation resulted in a 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. The specific amount of supplement was customized depending on each individual animal's body weight, the training phase it was in, and the motivational drive observed during the experiment.

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. The stimulus was presented typically immediately after the nose touch (delay = 0 ms), and mice were required to report its orientation with the appropriate nose-touch (Fig. 1B).

When stimulus size and contrast were manipulated (Fig. 1 and Fig. 2), the spatial frequency of the grating was fixed at 24 pixels/cycle, and three different sizes were tested: 60 x 60, 84 x 84, 108 x 108 (pixels x pixels). Seven different levels of contrast were tested in each case: $\text{luminance}_{\text{bright}}/\text{luminance}_{\text{dark}}$ = 1.5, 2.0, 3.3, 5.7, 12, 26, 99. Trials with different stimulus contrasts at a particular size were interleaved randomly throughout a session, while trials with different stimulus sizes were examined on different days. Data were recorded from a total of 18 sessions (days).

When stimulus size was manipulated independently (Fig.S1F-H), the spatial frequency and contrast of the grating were fixed, respectively, at 24 pixels/cycle and full contrast (99). Five different grating sizes were tested: 60 x 60, 84 x 84, 108 x 108, 132 x 132, 156 x 156 (pixels x pixels). Trials with different stimulus sizes were interleaved randomly throughout a session, and data were recorded from a total of five sessions (days).

When the stimulus duration was manipulated (Fig. 3), the spatial frequency, contrast, and size of the grating were fixed, respectively, at 24 pixels/cycle, full contrast (99), and 60 x 60 pixels x pixels. Eleven different stimulus durations were tested: 100, 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 over the range (from 3000 ms to 100 ms). Data were recorded from a total of 21 sessions.

When the stimulus onset delay, i.e., time difference between trial initiation and stimulus presentation, was manipulated (Fig. 6), the spatial frequency, contrast, size, and duration of the grating were fixed, respectively, at 24 pixels/cycle, full contrast (99), 60 x 60 pixels x pixels, and 600 ms. 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 over the range (from 0 ms to 200 ms). Data were recorded from a total of 7 sessions (days).

Flanker task. Upon trial initiation, either one stimulus ('target', 60 x 60 pixels, 2.5 cycle, 1s, contrast = 15.2) 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 and spatial frequency as the target, but of contrast in 8 different levels: 1.3, 1.7, 2.4, 3.8, 6.8, 15, 33, 99. 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. All types of trials (no flanker, congruent, incongruent) and flanker contrasts were interleaved randomly within each daily session. To train mice on this flanker task, they were first trained on the single stimulus discrimination task (with the target always at the lower location), following which, a flanker was introduced at the upper location with progressively increasing contrast over training days. Data from this experiment have been reported previously [16]. Those same data from the flanker task were re-analyzed here using different analyses, after collapsing trials across all the contrasts of the flanker.

Visual search task. Upon trial initiation, either one or two gratings were presented simultaneously (size= 60 x 60 pixels; duration = 3 s, delay = 0 ms; Fig.5A). When only one grating was presented (i.e., single target trial), it was presented above the central aperture, at the same horizontal level as the left and right

apertures, and mice were rewarded for reporting the orientation of the target as in the single discrimination task. When two gratings were presented simultaneously (i.e., search trial), one was presented just above the central aperture (center of grating was 30 pixels or 12.5º above the center of the central aperture; 'lower' location), and the other was presented far above the central aperture (center of the grating was 90 pixels or 37.5º above the center of the central aperture; 'upper' location). The orientations of the two gratings were always orthogonal to each other, their contrasts were always different from each other, and their relative contrast was systematically varied (Fig.5A). Mice were rewarded for reporting the orientation of the grating with higher contrast (i.e., 'target'), following the same rule as in the single stimulus conditions: vertical → nose touch to the left; horizontal → nose touch to the right. The target could appear at the upper or lower location (in a randomized fashion). Selection trials were interleaved with single target trials randomly throughout a session, and a total of 7 sessions (days) were recorded.

Subject inclusion/exclusion. A total of 33 mice were used. All 33 were trained on the single stimulus discrimination task, and of them, 8 mice never passed the inclusion threshold of % correct >70%. Of the remaining 25 mice, different subsets of mice were used to explore the limits of visual discrimination, and to study target selection. In general, for each task on which mice were trained (including the flanker and search tasks), they were considered to have learned the task if the overall performance (across all trial types) was > 70%. In the regular version of the search task (Fig. 5), all mice achieved this criterion. However, for the more challenging version of the search task (Fig. S3HI), only 2 /7 mice achieved this criterion.

For mice involved in more than one experiment, 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 response aperture within which to nose-touch. Mice used in visual search task were never used in the flanker task (and vice versa) since the two tasks involve different responding rules when two stimuli were presented simultaneously. In both the flaker and search tasks, we

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 response nose-touch, both detected by the touchscreen. Only in the case of the experiment involving stimulus onset delays (Fig. 6), another kind of 'reaction time' was measured for comparison. Denoted 'RT-start', this was the time of the response nose-touch calculated from the start of the trial, as opposed to from stimulus onset.

Conditional accuracy analysis. In order to get the full distribution of RT, trials from all mice were pooled together and treated as if they were from one single mouse. Pooled trials were then sorted by their RT, and then binned by RT into 100 ms or 200 ms bins, depending on the total number of trials available in each experiment. Conditional accuracy was calculated as the number of correct trials divided by the total number of trials for each RT bin.

Conditional accuracy function (CAF). To quantitatively describe the relationship between the conditional accuracy and RT, we fitted the plot of discrimination accuracy against (binned) RT with different functions (the CAF, see below) using a nonlinear least square method.

For RT bins aligned to stimulus onset (Fig. 2, 4C, S3, S4D), we fit the data using an asymptotic function: accuracy= $\lambda\ (1 - e^{-\gamma_{enc}\,(RT-\delta)})$). Three key metrics were defined for the sensory encoding phase for the use in subsequent comparisons between trial conditions: (1) peak conditional accuracy (a_{peak}), the maximal level of accuracy that the CAF reaches within the range of RT; (2) the timepoint at which the conditional accuracy reaches its maximal (t_{peak}). We defined it as the time point when the ascending CAF reaches a_{peak} *0.95; and (3) the timepoint at which the conditional accuracy just exceeds chance level of performance (t_{50}). We defined it as the time point when the ascending CAF crosses 52.5% (i.e., 50%*1.05). Note that t_{peak} and t_{50} are influenced by the slope parameter, γ_{enc}, and the temporal offset at chance performance, δ. Across different tasks, the parameter that was most commonly associated with

changes in the sensory encoding regime was a_{peak}; t_{peak} and t_{50} were largely unchanged by stimulus and task manipulations. It is possible that the lack of significant effects on t_{peak} and t_{50} was due to the very small proportion of trials with <300 ms RT across tasks (e.g., 2.4% in Fig. 2A), which resulted in higher variability in the estimates of t_{50} and t_{peak}.

For RT bins aligned to stimulus offset (Fig. 3, 4E, S4F), we fit the data using a sigmoid function: accuracy= $\lambda\ [1/(1 + e^{-\beta dec\ (RT- \tau)})]+50$ to quantify the time course of performance decay. Two key metrics were defined for this VSTM phase for the use in subsequent comparisons between trial conditions: (1) the first time point at which the conditional accuracy drops from its maximum (t_{decay}). We defined it as the time point when the descending CAF crosses a_{peak} *0.95; and (2) the first timepoint at which the conditional accuracy drops to a level indistinguishable from the chance (t_{chance}). We defined it as the timepoint when the descending CAF crosses 52.5%. In (rare) cases when the CAF never went below 52.5%, t_{chance} was set to be 3000ms. Note that t_{decay} and t_{chance} are influenced by the slope parameter, β_{dec}, and τ.

The confidence interval of the CAF and each metric were estimated by bootstrapping: the same number of trials were resampled from the raw data randomly with replacement, and were then processed following the same steps as described above to get repeated estimates of the CAF and corresponding metrics. Such resampling was repeated 1000 times to estimate the dispersion of each metric. Plots of the estimated value of each metric show the mean +/- std of the bootstrapped distribution of estimates (Fig.2C, 3E, 4DF, S5EG, S6EG).

Signal detection analysis (sensitivity and criterion). In the framework of signal detection theory, we assigned the correct vertical trials as 'hits', incorrect vertical trials as 'misses', correct horizontal trials as 'correct rejections' and incorrect horizontal trials as 'false alarms', and calculated the perceptual sensitivity (d') and criterion (c) accordingly [79] (Fig. S1B). Because of the inherent symmetry in 2-AFC tasks, this calculation was independent of which grating orientation – vertical or horizontal – was assigned as 'signal' and which as 'noise'. Consequently, a positive value of c caused poor performance just as much as the corresponding negative value, and therefore, we quantified the absolute value of c (|c|) as the relevant metric of decision criterion.

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 previously published, objective procedure [16], 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 performance in each bin (% correct), and plotted it as a function of trial number within session (Fig. S1EH, S2C, S3C, S4C). 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 (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. S1EH, S2C, S3C, S4C, along with the identified cut-off trial numbers (q).

Drift diffusion modeling of RT distributions. To shed light on potential mechanisms underlying observed RT distributions, we applied the drift-diffusion model to our RT data [48]. 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) during the viewing of the stimulus. A decision is said to be made when the accumulating evidence reaches an (abstract) internal

948 threshold of the subject. This process of evidence accumulation, together with the processes of sensory
949 encoding and motor execution, as well as threshold crossing, are said to determine the RT observed on each
950 trial.
951 We used a standard version of the model that consists of four independent variables [45-47]: (1) the drift
952 rate, (2) the boundary separation, (3) the starting point, and a (4) non-decisional constant, which accounts
953 for the time spent in sensory encoding and motor execution. In the case of our tasks, there was no reason
954 for the drift rate to be different between vertical versus horizontal gratings, and therefore, we merged both
955 type of trials (trials with a horizontal target grating and trials with a vertical target grating). We treated
956 'correct' response and 'incorrect' response as the two binary options, and fit the diffusion model to the RT
957 distributions of correct versus incorrect trials using the fast-dm-30 toolbox with the maximum likelihood
958 option to gain estimates of those four parameters for each individual mouse [48].
959 For accurate parameter estimates, trials with outlier values of RTs (too fast or too slow trials) are typically
960 excluded [48]. We identified inordinately fast or slow trials using a previously published procedure [16].
961 Briefly, for trials with RTs that are so short as to not allow mice sufficient time to accumulate sensory
962 evidence, performance would be consistently poor because mice would be forced to guess. Similarly, on
963 trials with RTs that are so long (far exceeding stimulus offset) as to extinguish the trace of sensory evidence
964 from their short-term memory [35, 46, 80], performance would be consistently poor because animals would
965 be forced to guess. From pooled RT data across all mice and all sessions that was binned in 50 ms bins, we
966 calculated the response accuracy and the 95% confidence intervals (using a bootstrapping method) for each
967 bin. Using this, we identified short and long RT bins for which the response accuracy was statistically
968 indistinguishable from chance and excluded them from the analysis.
969
970
971 **Statistical tests**: All analyses and statistical tests were performed in MATLAB. For single-stimulus
972 experiments in which only one stimulus parameter was systemically varied, one-way ANOVA was applied
973 to examine the effect of the manipulating the single factor (duration and delay, Fig. 3AB, 5C, 6, 7C, S1FG,
974 S2AB, S3A, S4AB, S5I-K). For experiments that involved changing both stimulus size and contrast
975 (Fig.1CD, 7A, S1CD, S5A-C) or changing both flanker congruency and contrast (Fig. 7B, S5E-G), two-
976 way ANOVA was applied to examine the effect of each factor, as well as their interaction.
977 For the flanker task, the *Wilcoxon signed-rank test* was used to examine if the group performance in each
978 trial type was different from chance, and also if there was difference between trial types (Fig.4B).
979 For the visual search task, the *Wilcoxon signed-rank test* was used to examine if the group performance
980 in each trial type was different from chance, and also if there was difference between trial types (Fig.5B).
981 One-way ANOVA was used to examine the effect of relative contrast (target/foil; Fig.5C). For comparisons
982 between single-stimulus trials and selection trials (Fig.5D), the signed-rank test was used when single-
983 stimulus were being compared to selection trials (with just one foil contrast), and the Kruskal-Wallis test
984 was used when trials from more than one foil contrast were being compared.
985 Correction for multiple comparisons was performed where necessary using the Holm-Bonferroni test
986 (HB test) for multiple comparisons.
987 The Pearson correlation coefficient and associated p-value were calculated for paired data (Fig. 2A, 3AB,
988 6BD, S1G, S5DHL) using *corrcoef* function in MATLAB.
989 For the metrics associated with CAF, permutation tests were used to determine if the estimated values
990 of each metric were different between experimental conditions (Fig.2CE, 3E, 4DF, S3EG, S4EG).
991 Specifically, trials from both conditions were pooled together (unlabeled), and randomly re-assigned into
992 two groups. The best-fit CAF and associated metrics were then calculated for each group, and so was the
993 difference of metrics between groups. Following 1000 repetitions, the resulting distribution of the
994 difference of metrics between groups was obtained under the null hypothesis that the data from the two
995 conditions were indistinguishable (i.e., from the same distribution). The real, observed difference of
996 metrics obtained from the best-fit CAF between two experimental conditions (for instance, Δa_{peak} from
997 low-contrast vs. high contrast conditions) was compared against the null distribution to compute the

corresponding p-value. Such p-values were corrected for multiple comparisons using the Holm-Bonferroni correction when multiple pairs of conditions were compared (Fig. 2C).

SUPPLEMENTARY INFORMATION

Supplementary figures (Fig. S1-S5) and legends are included below.

AUTHOR CONTRIBUTIONS

SPM and WKY designed the research and wrote the paper. WKY performed the experiments, and analyzed the data.

ACKNOWLEDGEMENTS

This work was supported by funding from NIH R03 HD093995 and startup funds from Johns Hopkins University. We thank James Garmon for help with fabrication of custom equipment.

COMPETING FINANCIAL INTERESTS

The authors declare that there are no competing financial interests.

REFERENCES

1. Huberman, A.D. and C.M. Niell, *What can mice tell us about how vision works?* Trends Neurosci, 2011. **34**(9): p. 464-73.
2. Glickfeld, L.L., R.C. Reid, and M.L. Andermann, *A mouse model of higher visual cortical function.* Current opinion in neurobiology, 2014. **24**: p. 28-33.
3. Seabrook, T.A., et al., *Architecture, function, and assembly of the mouse visual system.* Annual review of neuroscience, 2017. **40**: p. 499-538.
4. Burgess, C.P., et al., *High-Yield Methods for Accurate Two-Alternative Visual Psychophysics in Head-Fixed Mice.* Cell Rep, 2017. **20**(10): p. 2513-2524.
5. Marques, T., et al., *A Role for Mouse Primary Visual Cortex in Motion Perception.* Curr Biol, 2018. **28**(11): p. 1703-1713.e6.
6. Busse, L., et al., *The detection of visual contrast in the behaving mouse.* J Neurosci, 2011. **31**(31): p. 11351-61.
7. Carandini, M. and A.K. Churchland, *Probing perceptual decisions in rodents.* Nature neuroscience, 2013. **16**(7): p. 824.
8. Glickfeld, L.L., M.H. Histed, and J.H. Maunsell, *Mouse primary visual cortex is used to detect both orientation and contrast changes.* J Neurosci, 2013. **33**(50): p. 19416-22.
9. Long, M., et al., *Contrast-dependent orientation discrimination in the mouse.* Sci Rep, 2015. **5**: p. 15830.
10. Histed, M.H., L.A. Carvalho, and J.H. Maunsell, *Psychophysical measurement of contrast sensitivity in the behaving mouse.* J Neurophysiol, 2012. **107**(3): p. 758-65.
11. Prusky, G.T. and R.M. Douglas, *Characterization of mouse cortical spatial vision.* Vision Res, 2004. **44**(28): p. 3411-8.
12. Prusky, G.T., P.W. West, and R.M. Douglas, *Behavioral assessment of visual acuity in mice and rats.* Vision Res, 2000. **40**(16): p. 2201-9.
13. Dent, C.L. and A.R. Isles, *An overview of measuring impulsive behavior in mice.* Curr Protoc Mouse Biol, 2014. **4**(2): p. 35-45.
14. Isles, A.R., et al., *Common genetic effects on variation in impulsivity and activity in mice.* J Neurosci, 2004. **24**(30): p. 6733-40.
15. Wang, L. and R.J. Krauzlis, *Visual Selective Attention in Mice.* Curr Biol, 2018. **28**(5): p. 676-685.e4.
16. You, W.-K. and S.P. Mysore, *Endogenous and exogenous control of visuospatial selective attention in freely behaving mice.* bioRxiv, 2019: p. 550822.
17. Mar, A.C., et al., *The touchscreen operant platform for assessing executive function in rats and mice.* Nat Protoc, 2013. **8**(10): p. 1985-2005.
18. Wickelgren, W.A., *Speed-accuracy tradeoff and information processing dynamics.* Acta psychologica, 1977. **41**(1): p. 67-85.

19. Luce, R.D., *Response times: Their role in inferring elementary mental organization*. 1986: Oxford University Press on Demand.
20. McElree, B. and B.A. Dosher, *Serial position and set size in short-term memory: the time course of recognition.* Journal of Experimental Psychology: General, 1989. **118**(4): p. 346.
21. Heitz, R.P., *The speed-accuracy tradeoff: history, physiology, methodology, and behavior.* Frontiers in neuroscience, 2014. **8**: p. 150.
22. Shibuya, H. and C. Bundesen, *Visual selection from multielement displays: measuring and modeling effects of exposure duration.* Journal of Experimental Psychology: Human Perception and Performance, 1988. **14**(4): p. 591.
23. Watamaniuk, S.N. and R. Sekuler, *Temporal and spatial integration in dynamic random-dot stimuli.* Vision research, 1992. **32**(12): p. 2341-2347.
24. Busey, T.A. and G.R. Loftus, *Sensory and cognitive components of visual information acquisition.* Psychological Review, 1994. **101**(3): p. 446.
25. Vogel, E.K., G.F. Woodman, and S.J. Luck, *The time course of consolidation in visual working memory.* Journal of Experimental Psychology: Human Perception and Performance, 2006. **32**(6): p. 1436.
26. Bays, P.M., et al., *Temporal dynamics of encoding, storage, and reallocation of visual working memory.* Journal of vision, 2011. **11**(10): p. 6-6.
27. Sperling, G., *The information available in brief visual presentations.* Psychological monographs: General and applied, 1960. **74**(11): p. 1.
28. Averbach, E. and A.S. Coriell, *Short-term memory in vision.* The Bell System Technical Journal, 1961. **40**(1): p. 309-328.
29. Posner, M.I. and S.W. Keele, *Decay of visual information from a single letter.* Science, 1967. **158**(3797): p. 137-139.
30. Phillips, W. and A. Baddeley, *Reaction time and short-term visual memory*. Psychonomic Science, 1971. **22**(2): p. 73-74.
31. Phillips, W., *On the distinction between sensory storage and short-term visual memory*. Perception & Psychophysics, 1974. **16**(2): p. 283-290.
32. Dick, A., *Iconic memory and its relation to perceptual processing and other memory mechanisms.* Perception & Psychophysics, 1974. **16**(3): p. 575-596.
33. Coltheart, M., *Iconic memory and visible persistence.* Perception & psychophysics, 1980. **27**(3): p. 183-228.
34. Magnussen, S. and M.W. Greenlee, *The psychophysics of perceptual memory*. Psychological research, 1999. **62**(2-3): p. 81-92.
35. Smith, P.L. and R. Ratcliff, *An integrated theory of attention and decision making in visual signal detection.* Psychol Rev, 2009. **116**(2): p. 283-317.
36. Brown, J., *Some tests of the decay theory of immediate memory*. Quarterly Journal of Experimental Psychology, 1958. **10**(1): p. 12-21.
37. Gold, J.M., et al., *Visual memory decay is deterministic.* Psychological Science, 2005. **16**(10): p. 769-774.
38. Zhang, W. and S.J. Luck, *Sudden death and gradual decay in visual working memory.* Psychological science, 2009. **20**(4): p. 423-428.
39. Barrouillet, P. and V. Camos, *As time goes by: Temporal constraints in working memory*. Current Directions in Psychological Science, 2012. **21**(6): p. 413-419.
40. Ricker, T.J., E. Vergauwe, and N. Cowan, *Decay theory of immediate memory: From Brown (1958) to today (2014).* The Quarterly Journal of Experimental Psychology, 2016. **69**(10): p. 1969-1995.
41. Di Lollo, V., *Temporal characteristics of iconic memory.* Nature, 1977. **267**(5608): p. 241-243.
42. Di Lollo, V., *Temporal integration in visual memory*. Journal of Experimental Psychology: General, 1980. **109**(1): p. 75.
43. Eriksen, B.A. and C.W. Eriksen, *Effects of noise letters upon the identification of a target letter in a nonsearch task.* Perception & Psychophysics, 1974. **16**(1): p. 143-149.
44. Fan, J., et al., *Testing the efficiency and independence of attentional networks.* J Cogn Neurosci, 2002. **14**(3): p. 340-7.
45. Ratcliff, R., *A theory of memory retrieval.* Psychological Review, 1978. **85**(2): p. 59-108.
46. Ratcliff, R., et al., *Diffusion Decision Model: Current Issues and History.* Trends Cogn Sci, 2016. **20**(4): p. 260-281.
47. Voss, A., M. Nagler, and V. Lerche, *Diffusion models in experimental psychology: a practical introduction.* Exp Psychol, 2013. **60**(6): p. 385-402.

48. Voss, A., J. Voss, and V. Lerche, *Assessing cognitive processes with diffusion model analyses: a tutorial based on fast-dm-30.* Front Psychol, 2015. **6**: p. 336.
49. Wong, A.A. and R.E. Brown, *Visual detection, pattern discrimination and visual acuity in 14 strains of mice.* Genes Brain Behav, 2006. **5**(5): p. 389-403.
50. Ohzawa, I., G. Sclar, and R.D. Freeman, *Contrast gain control in the cat visual cortex.* Nature, 1982. **298**(5871): p. 266-8.
51. Roitman, J.D. and M.N. Shadlen, *Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task.* Journal of neuroscience, 2002. **22**(21): p. 9475-9489.
52. Zylberberg, A., C.R. Fetsch, and M.N. Shadlen, *The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision.* Elife, 2016. **5**: p. e17688.
53. Uchida, N. and Z.F. Mainen, *Speed and accuracy of olfactory discrimination in the rat.* Nat Neurosci, 2003. **6**(11): p. 1224-9.
54. Zariwala, H.A., et al., *The limits of deliberation in a perceptual decision task.* Neuron, 2013. **78**(2): p. 339-51.
55. Uchida, N., A. Kepecs, and Z.F. Mainen, *Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making.* Nature Reviews Neuroscience, 2006. **7**(6): p. 485-491.
56. Abraham, N.M., et al., *Maintaining accuracy at the expense of speed: stimulus similarity defines odor discrimination time in mice.* Neuron, 2004. **44**(5): p. 865-876.
57. Khan, R.M. and N. Sobel, *Neural processing at the speed of smell.* Neuron, 2004. **44**(5): p. 744-747.
58. Neisser, U., *Cognitive psychology*. Cognitive psychology. 1967, East Norwalk, CT, US: Appleton-Century-Crofts.
59. Sperling, G., *A model for visual memory tasks.* Human factors, 1963. **5**(1): p. 19-31.
60. Eriksen, C.W. and J.F. Collins, *Sensory traces versus the psychological moment in the temporal organization of form.* Journal of Experimental Psychology, 1968. **77**(3p1): p. 376.
61. Holding, D.H., *Sensory storage reconsidered.* Memory & Cognition, 1975. **3**(1): p. 31-41.
62. Fougnie, D. and G.A. Alvarez, *Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model.* Journal of vision, 2011. **11**(12): p. 3-3.
63. Haber, R.N. and M. Hershenson, *The psychology of visual perception*. 1973: Holt, Rinehart & Winston.
64. Potter, M.C., *Short-term conceptual memory for pictures.* Journal of Experimental Psychology: Human Learning and Memory, 1976. **2**(5): p. 509-522.
65. Raymond, J.E., K.L. Shapiro, and K.M. Arnell, *Temporary suppression of visual processing in an RSVP task: An attentional blink?* Journal of experimental psychology: Human perception and performance, 1992. **18**(3): p. 849.
66. Duncan, J., R. Ward, and K. Shapiro, *Direct measurement of attentional dwell time in human vision.* Nature, 1994. **369**(6478): p. 313-315.
67. Chun, M.M. and M.C. Potter, *A two-stage model for multiple target detection in rapid serial visual presentation.* Journal of Experimental Psychology: Human Perception and Performance, 1995. **21**(1): p. 109-127.
68. Ward, R., J. Duncan, and K. Shapiro, *The slow time-course of visual attention.* Cognitive psychology, 1996. **30**(1): p. 79-109.
69. Jolicœur, P. and R. Dell'Acqua, *The demonstration of short-term consolidation.* Cognitive psychology, 1998. **36**(2): p. 138-202.
70. Vogel, E.K., S.J. Luck, and K.L. Shapiro, *Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink.* Journal of Experimental Psychology: Human Perception and Performance, 1998. **24**(6): p. 1656.
71. Umino, Y., R. Pasquale, and E. Solessio, *Visual temporal contrast sensitivity in the behaving mouse shares fundamental properties with human psychophysics.* eNeuro, 2018. **5**(4).
72. Treisman, A. and S. Gormican, *Feature analysis in early vision: evidence from search asymmetries.* Psychological review, 1988. **95**(1): p. 15.
73. Wolfe, J.M., *Visual search*, in *Attention*. 1998, Psychology Press/Erlbaum (UK) Taylor & Francis: Hove, England. p. 13-73.
74. Yantis, S. and H.E. Egeth, *On the distinction between visual salience and stimulus-driven attentional capture.* J Exp Psychol Hum Percept Perform, 1999. **25**(3): p. 661-76.
75. Sridharan, D., et al., *Visuospatial selective attention in chickens.* Proc Natl Acad Sci U S A, 2014. **111**(19): p. E2056-65.

76. Chung, S.T., D.M. Levi, and G.E. Legge, *Spatial-frequency and contrast properties of crowding.* Vision Res, 2001. **41**(14): p. 1833-50.
77. Pelli, D.G., M. Palomares, and N.J. Majaj, *Crowding is unlike ordinary masking: distinguishing feature integration from detection.* J Vis, 2004. **4**(12): p. 1136-69.
78. Guo, Z.V., et al., *Procedures for behavioral experiments in head-fixed mice.* PLoS One, 2014. **9**(2): p. e88678.
79. Stanislaw, H. and N. Todorov, *Calculation of signal detection theory measures.* Behav Res Methods Instrum Comput, 1999. **31**(1): p. 137-49.
80. Ratcliff, R. and J.N. Rouder, *A diffusion model account of masking in two-choice letter identification.* J Exp Psychol Hum Percept Perform, 2000. **26**(1): p. 127-40.

Supplementary Information

Visual psychophysics and limits of visual discrimination performance in freely behaving mice

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

*Corresponding author: shreesh.mysore@jhu.edu

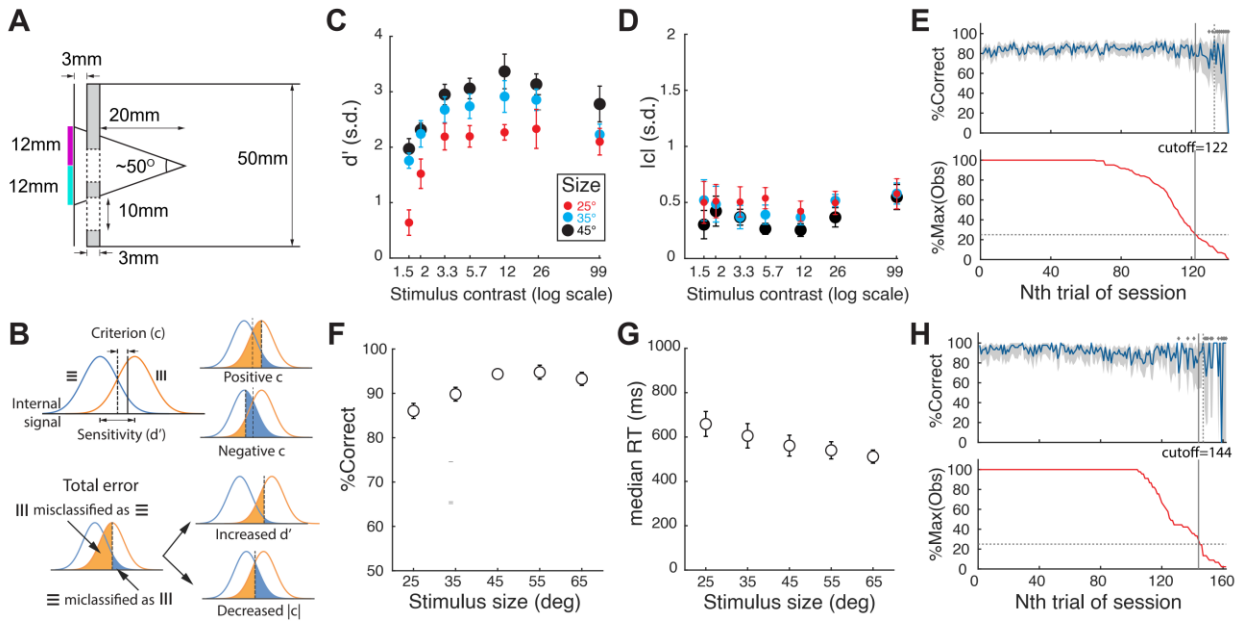

Figure S1. Stimulus contrast and size modulate perceptual sensitivity but not decision criterion of orientation discrimination performance. (A) Lateral view of the schematic experimental setup showing the relative position of the touchscreen (leftmost vertical line), the plexiglass mask (grey-filled 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.1B). For experiments involving two stimulus locations (i.e., flanker task and search task), the upper (magenta) and lower (cyan) locations of the stimulus are indicated as colored bars (see also Fig. 4A, 5A). 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)** Schematic of the signal detection theory (SDT) analysis illustrating perceptual sensitivity (d') and criterion (c) calculations for 2-AFC task (Methods). Upper row; left: SDT hypothesizes that the internal representation of vertical and horizontal stimuli can be reduced (projected) to a one-dimensional decision axis, on which they form two overlapping distributions (due to noise). A decision is made based on a criterion set by each individual animal: a stimulus whose representation falls above (or below) the criterion is judged as vertical (or horizontal), producing the appropriate behavioral response. A decision criterion (c) of 0, by definition, corresponds to optimal (unbiased) performance given the two distributions. For our 2-AFC task, we defined the decision criterion as the amount of deviation from an unbiased value for the following reason. Upper row; right: Because of the inherent symmetry of 2-AFC task design, positive criterion would increase errors in classification of vertical targets, but also slightly decrease errors in classification of horizontal targets, producing a net reduction in overall accuracy. Similarly, a negative criterion would increase errors in classification of horizontal targets, but also slightly decrease errors in classification of vertical targets, again producing a net decrease in overall accuracy. Therefore, when the two distributions are similar, a negative as well as a positive criterion of the same magnitude will produce the same overall reduction in discrimination accuracy, but a criterion of smaller absolute value would signal an overall improvement in performance. For this reason, we used the absolute value of c (|c|) to examine the effect of criterion change on response accuracy. Lower row: Based on theory, improved response accuracy can result from (1) increased d': when the two distributions become further separated; or (2) decreased |c|: when the decision criterion becomes less biased. **(C)** Plot of perceptual sensitivity against stimulus contrast ($luminance_{Bright}/luminance_{Dark}$; log scale; Methods). Different colors correspond to different dot sizes. Data: mean ± s.e.m; n= 8 mice. 2-way ANOVA, p<0.001 (contrast), p<0.001 (size), p=0.899 (interaction). **(D)** Plot of absolute value of criterion |c| against stimulus contrast (log scale). Data: mean ± s.e.m; n= 8 mice. 2-way ANOVA, p=0.374 (contrast), p=0.056 (size), p=0.998 (interaction). **(E)** Identification of trials towards the end of the 30 min behavioral sessions that corresponded to animals being poorly engaged in the task (Methods). 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. 1CD and S1CD are based on data from trials 1-121 from each behavioral session. **(F-H)** Behavioral response metrics as a function of stimulus size (n=9 mice). (F) Discrimination accuracy; p=0.001, 1-way ANOVA against stimulus size. (G) Median RT; p=0.205, 1-way ANOVA against stimulus size. Data: mean ± s.e.m; n= 9 mice. (H) 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 E.

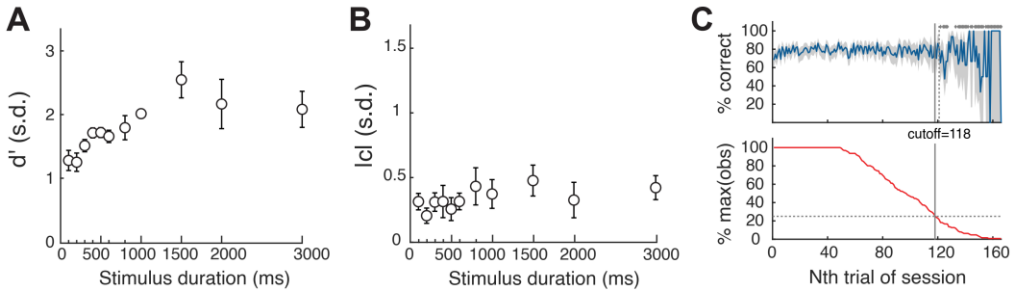

Figure S2. Stimulus duration modulates perceptual sensitivity but not decision criterion of orientation discrimination performance. (A) Plot of perceptual sensitivity against stimulus duration. Data: mean ± s.e.m; n= 6 mice. 1-way ANOVA, p=0.001. **(B)** Plot of absolute value of criterion |c| against stimulus duration. Data: mean ± s.e.m; n= 6 mice. 1-way ANOVA, p=0.802. **(C)** 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. S1E.

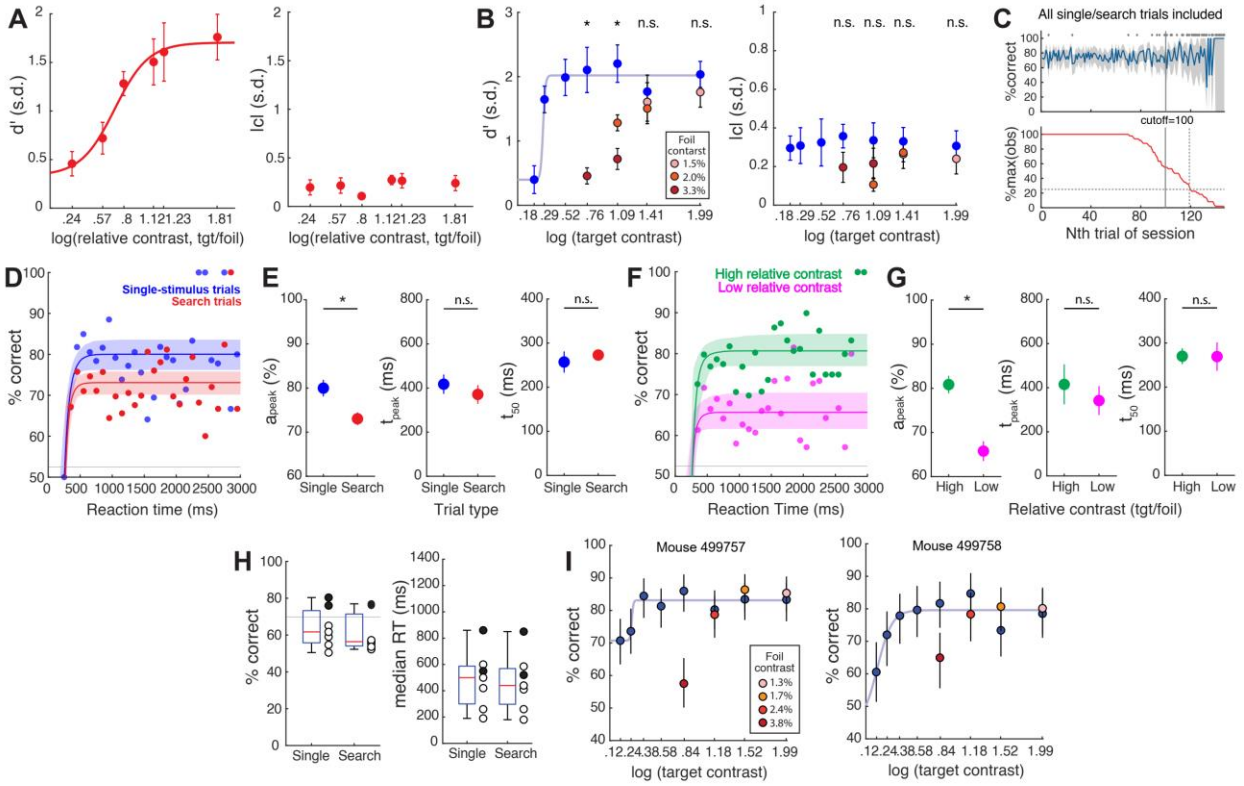

Figure S3. Foil modulates sensory encoding regime in rudimentary visual search task. (A) Perceptual sensitivity (d', left) and criterion (|c|, right) plotted against relative contrast of target:foil. Red curve: best sigmoidal fit (Methods). Left: p<0.001, 1-way ANOVA. Right: p=0.552, 1-way ANOVA. **(B)** Comparison of discrimination performance to target presented in the presence of foil (red data points) vs. when target was presented alone (blue data points). Left: Perceptual sensitivity (d'). Right: criterion (|c|). Shades of red: contrasts of foil. '*' ('n.s.'): p<0.05 (p>0.05), Kruskal-Wallis tests followed by HBMC correction (Methods). **(C)** 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. S1E. Based on these data, all trials above 100 of each behavioral session of this experiment were dropped from analysis. Results in Fig. 5 and S3 are based on data from trials 1-99 from each behavioral session. **(D)** CAFs for different trial types (blue: single-stimulus trials; red: search trials (Methods). **(E)** Plots of parameters of the CAF for different trial types. Single-stimulus vs. search trials: a_{peak}: p=0.004; t_{peak}: p=0.418; t_{50}: p=0.434, permutation tests (Methods). **(F)** CAFs for different relative contrast (target:foil) levels. Magenta: low relative contrast, first three relative contrast levels from Fig. 5C; Green: high relative contrast, the last three relative contrast levels from Fig. 5C (Methods). **(G)** Plots of parameters of the CAF for different relative contrast levels. High vs. low relative contrast: a_{peak}: p<0.001; t_{peak}: p=0.349; t_{50}: p=0.944, permutation tests (Methods). **(H-I)** Performance of mice in the difficult version of search task (stimulus duration = 800 ms). (H) Left panel: response accuracy; right panel: median RT; data from trials pooled across all relative contrast conditions. Only two mice (filled black circles) passed the inclusion criterion with overall response accuracy >70%. (I) Performance of the two mice (left and right panels, respectively) that successfully learned the difficult version of the search task. Shown is the comparison of discrimination performance to target presented in the presence of foil (red data points) vs. when target was presented alone (blue data points) of those two mice in the advanced search task. The presence of foil reduces target discrimination accuracy in a contrast-dependent manner –a pattern similar/identical to that in Fig. 5D. Shades of red: contrasts of foil. Error bars: 95% C.I. estimated via bootstrapping.

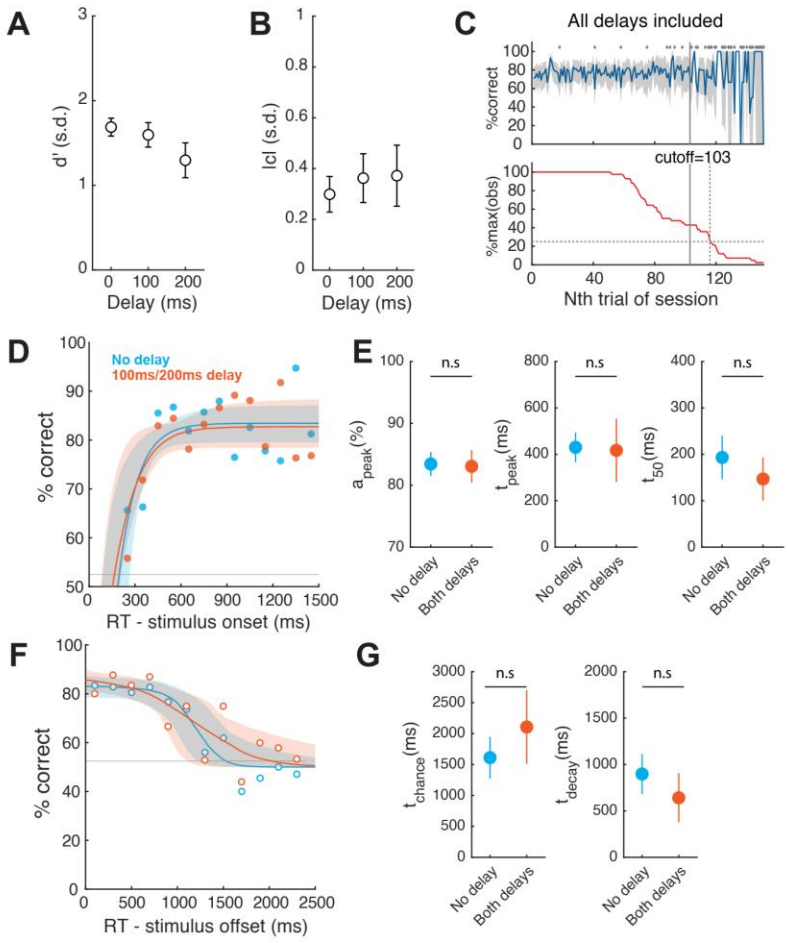

Figure S4. Stimulus onset delay modulates neither the sensory encoding nor the VSTM regimes of the CAF. **(A)** Plot of perceptual sensitivity (d') against stimulus delay; p=0.217, 1-way ANOVA. **(B)** Plot of criterion (|c|) against stimulus delay; p=0.848, 1-way ANOVA. **(C)** Identification of trials 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. S1E. Based on these data, all trials above 103 of each behavioral session of this experiment were dropped from analysis. Results in Fig. 6 and S4 are based on data from trials 1-102 from each behavioral session. **(D)** CAFs of the sensory encoding stage, for targets of different stimulus onset delays; data correspond to trials with RT < 1500 ms. Blue: trials with no delay; orange: trials with 100 ms or 200 ms onset delays (Methods). **(E)** Plots of key parameters of CAFs in D (sensory encoding regime) for different trial types. Data show mean ± s.t.d of distribution of bootstrapped estimates. Delay vs. no delay: a_{peak}: p=0.921; t_{peak}: p=0.887; t_{50}: p=0.105, permutation tests (Methods). **(F)** CAFs of the VSTM-dependent stage, for targets of different stimulus onset delays; data correspond to trials with RT > stimulus offset (600 ms), aligned to stimulus offset. Blue: no delay; orange: with 100 ms or 200 ms stimulus onset delay (Methods). **(G)** Plots of parameters of the CAF in (F) for different trial types. Delay vs. no delay: t_{chance}: p=0.064; t_{peak}: p=0.156, permutation tests (Methods).

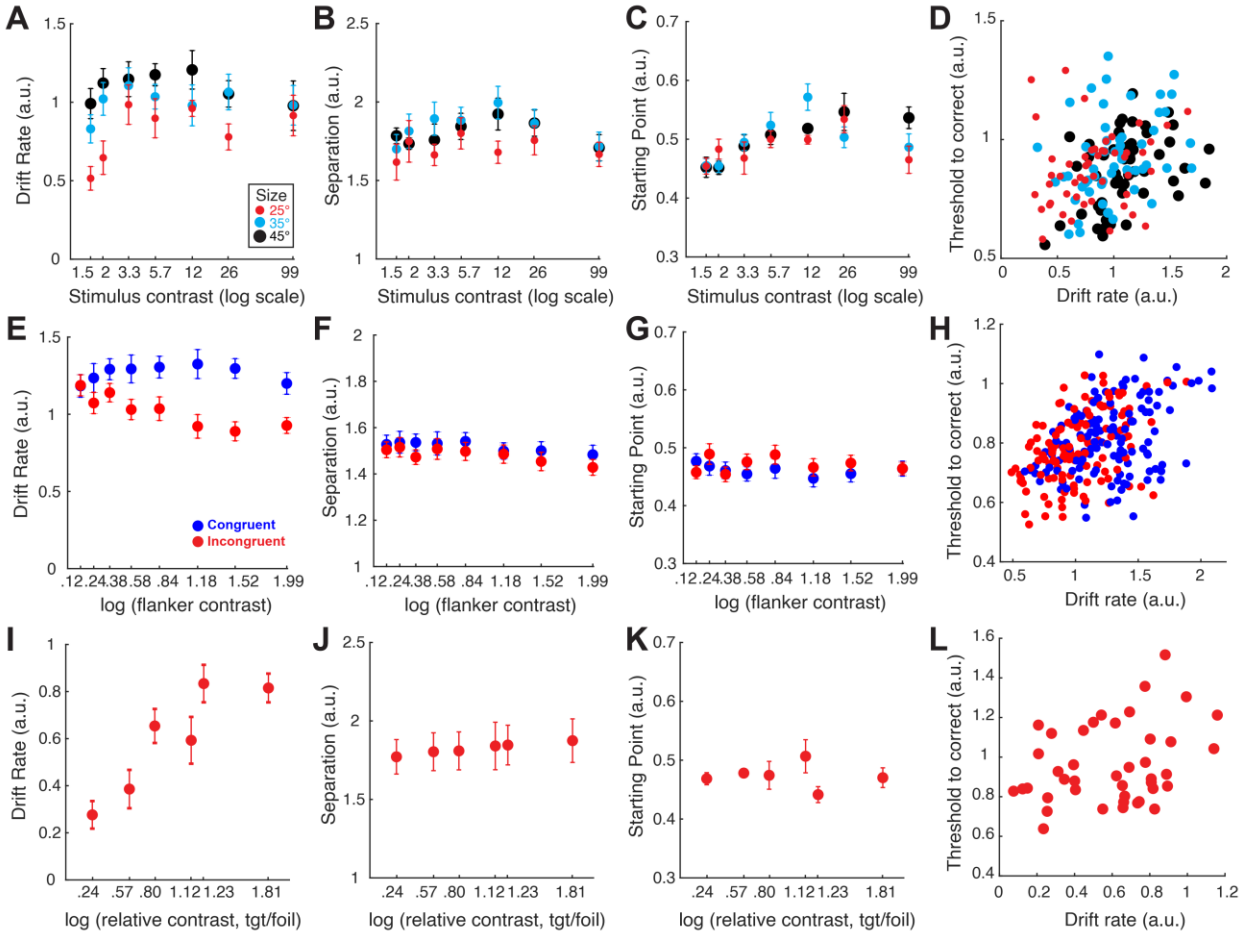

Figure S5. Estimates of parameters of the drift diffusion model applied to data from different tasks. **(A-D)** Single-stimulus discrimination task. Each model parameter plotted against stimulus size and contrast, as in Fig. 1. (A) drift rate: 2-way ANOVA, p=0.028 (contrast), p<0.001 (size), p=0.767 (interaction); (B) boundary separation: 2-way ANOVA, p=0.171 (contrast), p=0.026 (size), p=0.953 (interaction); (C) starting point: 2-way ANOVA, p<0.001 (contrast), p=0.325 (size), p=0.098 (interaction). (D) Scatter plot of threshold to correct response versus drift rate; each dot is one individual mouse; colors as in A; Pearson's correlation = 0.3, p<0.001. **(E-H)** Flanker task. Each model parameter plotted against flanker congruency and contrast. (E) drift rate: 2-way ANOVA, p<0.001 (flanker congruency), p=0.475 (flanker contrast), p=0.097 (interaction); (F) boundary separation: 2-way ANOVA, p=0.069 (flanker congruency), p=0.617 (flanker contrast), p=0.998 (interaction); (G) starting point: 2-way ANOVA, p=0.173 (flanker congruency), p=0.741 (flanker contrast), p=0.724 (interaction). (H) Scatter plot of threshold to correct response versus drift rate; each dot is one individual mouse; colors as in E; Pearson's correlation = 0.42, p<0.001. **(I-L)** Search task. Each model parameter plotted against the relative contrast of target:foil. (I) drift rate: p<0.001, 1-way ANOVA; (J) boundary separation: p=0.997, 1-way ANOVA; (K) starting point: p=0.275, 1-way ANOVA. (L) Scatter plot of threshold to correct response versus drift rate; each dot is one individual mouse; Pearson's correlation = 0.3, p=0.05.