

Reviewed Preprint

v1 • April 1, 2026

Not revised

Reviewed Preprint

v2 • July 13, 2026

Revised by authors

✉ For correspondence:

jigold@pennteamedicine.upenn.edu

Competing interests:

No competing interests declared

Funding:

See [page 29](#)

Reviewing editor:

Redmond G O'Connell, Trinity College Dublin, Ireland

© 2026, McGaughey & Gold. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Sensory adaptation and pupil-linked arousal support flexible evidence accumulation during perceptual decision making

Kara D McGaughey^{1,2,3}, Joshua I Gold^{1,2,3} ✉

Department of Neuroscience, University of Pennsylvania, Philadelphia, United States • Computational Neuroscience Initiative, University of Pennsylvania, Philadelphia, United States • Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, United States

eLife Assessment

This **important** study measures single-unit activity in area MT of awake-behaving monkeys to test the idea that sensory adaptation contributes to flexible evidence accumulation during decision making. The authors provide **compelling** evidence that adaptation to different temporal contexts shapes both perceptual judgements and neural responses. Although the precise computational mechanisms underlying these effects remain uncertain, the results support the conclusion that recent sensory history influences the temporal dynamics of decision formation. This work will be of interest to researchers studying visual perception, sensory adaptation, and decision making.

<https://doi.org/10.7554/eLife.110685.2.sa3>

Abstract

Effective decision making in dynamic environments requires flexible evidence accumulation. Although models often express this flexibility as a property of the accumulator, its implementation in the brain may involve adaptive mechanisms operating at other stages of the decision process. We examined two such mechanisms: 1) stimulus-specific sensory adaptation at the level of evidence encoding; and 2) arousal-related neuromodulation, which could, in principle, affect both evidence encoding and accumulation. We measured single-unit activity in the middle temporal area (MT) and pupil-linked arousal while monkeys performed a modified random-dot motion direction-discrimination task in which an adapting stimulus with varied temporal stability preceded a behaviorally relevant test stimulus. The monkeys' decisions reflected adaptive evidence accumulation that depended on temporal-context stability and corresponded to context-dependent changes in both stimulus-specific sensory adaptation in MT and task-evoked pupil responses. However, adaptation and pupil adjustments were not related to each other. Together, these findings suggest that multiple mechanisms contribute to flexible, context-dependent evidence accumulation, including changes in sensory adaptation that shape evidence encoding and changes in arousal that may shape the accumulation process itself.

Introduction

Perceptual decisions often require the accumulation of uncertain sensory evidence over time. This accumulation process is complicated by the fact that the nature and source of incoming evidence can change even while accumulation is underway. Under these changing conditions, effective decision makers should adjust how sensory information is accumulated so that their perceptual judgments reflect current conditions rather than outdated evidence gathered before a change occurred (Gold & Stocker, 2017 [↗](#)). This flexibility can be achieved by adjusting the evidence-

accumulation process in several ways, including adopting a leak and/or saturating non-linearity that can be adjusted according to environmental conditions to balance sensitivity to stability versus change (Glaze et al., 2015 [DOI](#); Usher & McClelland, 2001 [DOI](#)). These adaptive forms of evidence accumulation are key features of decision-making behavior in dynamic environments (Glaze et al., 2015 [DOI](#); Murphy et al., 2021 [DOI](#); Piet et al., 2018 [DOI](#)) and are often modeled using one or more parameters that control the temporal dynamics of an accumulator (Glaze et al., 2015 [DOI](#); Usher & McClelland, 2001 [DOI](#); Veliz-Cuba et al., 2016 [DOI](#), Fig. 1 [DOI](#)).

However, unlike these simple models, the brain contains many mechanisms that could, in principle, shape the temporal dynamics of evidence accumulation. Here, we considered a potential role for stimulus-specific sensory adaptation of individual neurons encoding the evidence that is accumulated downstream to form the decision (Fig. 1 [DOI](#)). Adaptation is a hallmark of sensory systems, occurring across modalities (Dalton, 2000 [DOI](#); Lampl & Katz, 2017 [DOI](#); Laughlin, 1989 [DOI](#); Ulanovsky et al., 2004 [DOI](#)) and species (Brenner et al., 2000 [DOI](#); Clarke et al., 2015 [DOI](#); Lesica et al., 2007 [DOI](#); Nagel & Doupe, 2006 [DOI](#); Tolias et al., 2001 [DOI](#)). Recent theoretical work has framed adaptation as a form of active inference, whereby sensory systems not only modulate their responses to match the statistical structure of incoming signals, but also infer when those statistics have changed and how rapidly they are expected to vary (DeWeese & Zador, 1998 [DOI](#); Młynarski & Hermundstad, 2021 [DOI](#); Solomon & Kohn, 2014 [DOI](#); Wark et al., 2007 [DOI](#); Weber et al., 2019 [DOI](#)). Consistent with this idea, recordings from isolated retinal ganglion cells have shown that the temporal dynamics of recent stimulus history can modulate adaptation (Wark et al., 2009 [DOI](#)). Yet it remains unknown whether individual neurons in primate sensory cortex undergo similar temporal context-dependent adaptation, and whether such adaptation contributes to flexible evidence accumulation for perceptual decision making.

To test these ideas, we leveraged decisions about visual motion. These decisions depend on neurons in the middle temporal area (MT) of extrastriate visual cortex (K. Britten et al., 1992 [DOI](#); K. H. Britten et al., 1996 [DOI](#); Hanks et al., 2006 [DOI](#); Newsome & Pare, 1988 [DOI](#); Salzman et al., 1990 [DOI](#)), which exhibit robust stimulus-specific sensory adaptation (Kohn & Movshon, 2003 [DOI](#); Van Wezel & Britten, 2002 [DOI](#)). We manipulated the stability of recent sensory experience by modifying the classic random-dot motion direction-discrimination task to include extended exposure to an adapting stimulus that reversed direction at either a low or high frequency, creating two context-stability conditions (Fig. 2A [DOI](#)). Immediately following this adapting stimulus, monkeys viewed and reported the direction of a test stimulus, which was identical across conditions, allowing us to isolate the influence of recent stimulus statistics on both behavior and neural activity. We found that monkeys adjusted their decision-making behavior in a manner that was consistent with context-dependent changes in the temporal dynamics of evidence accumulation. These adjustments were based, in part, on differences in how the evidence being accumulated was encoded by individual MT neurons.

We also examined whether and how these context-dependent behavioral and neural adjustments related to simultaneous measures of pupil diameter. Non-luminance-mediated fluctuations in pupil diameter, an index of arousal, track quantities relevant to adaptive evidence accumulation, including environmental volatility and temporal expectations (Akdoğan et al., 2016 [DOI](#); Nassar et al., 2012 [DOI](#); Schwiedrzik & Sudmann, 2020 [DOI](#)). Pupil diameter also relates to perceptual sensitivity, bias, and variability in evidence accumulation (de Gee et al., 2014 [DOI](#), 2017 [DOI](#); Keung et al., 2019 [DOI](#); Krishnamurthy et al., 2017 [DOI](#); Murphy et al., 2014 [DOI](#)). These findings imply a role for the arousal system in optimizing perceptual decision making, likely via neuromodulatory effects on neural information-processing dynamics (Aston-Jones & Cohen, 2005 [DOI](#); Devilbiss et al., 2006 [DOI](#); Joshi & Gold, 2020 [DOI](#); Navarra et al., 2017 [DOI](#); Waterhouse et al., 1998 [DOI](#); Wyrick & Mazzucato, 2021 [DOI](#)). Arousal systems are therefore well positioned to, in principle, influence how sensory information is encoded, accumulated, or both (Fig. 1 [DOI](#)). Our results, which show a dissociation between sensory adaptation-related and arousal-related effects on evidence accumulation, provide new insights into the complex, distributed mechanisms that contribute to flexible decision making.

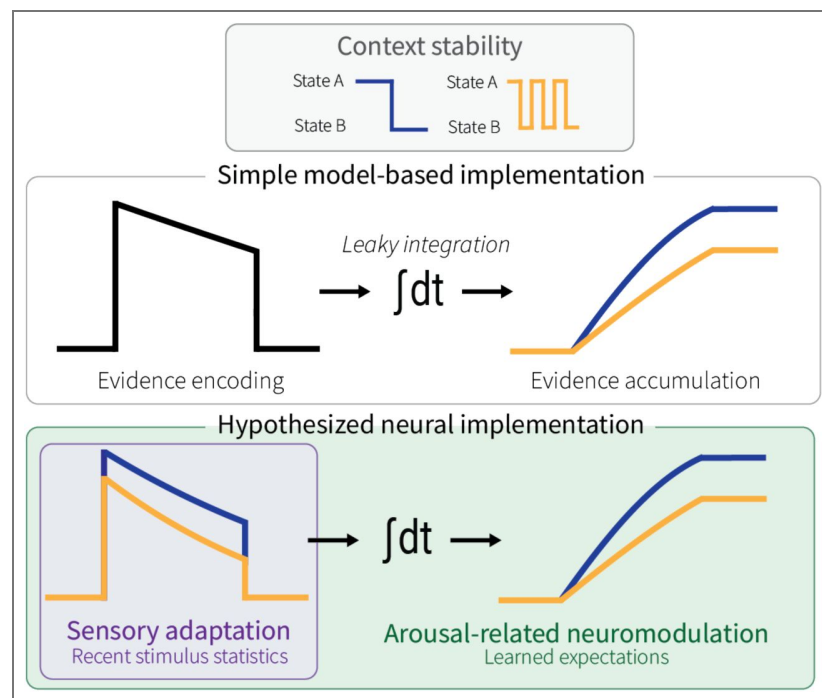


Fig. 1. Hypothesized neural implementation of context-dependent changes in evidence accumulation.

For perceptual decisions about motion direction, momentary evidence is encoded in the middle temporal area (MT) and then accumulated by downstream circuits to form a decision variable that guides behavior. Models of decision-making often implement flexible, context-dependent evidence accumulation via a single parameter that controls the temporal dynamics (e.g., leakiness) of an accumulator (*top*). In principle, adjustments to the accumulation process could involve changes in evidence encoding, accumulation, or both (*bottom*). Here we used manipulations of context stability (low-versus high-frequency direction reversals of an adapting motion stimulus) to test for a role of evidence encoding. Because global signals related to arousal can influence decision flexibility and cortical information processing at multiple levels, we further considered potential interactions between adaptation- and arousal-related effects on flexible evidence accumulation.

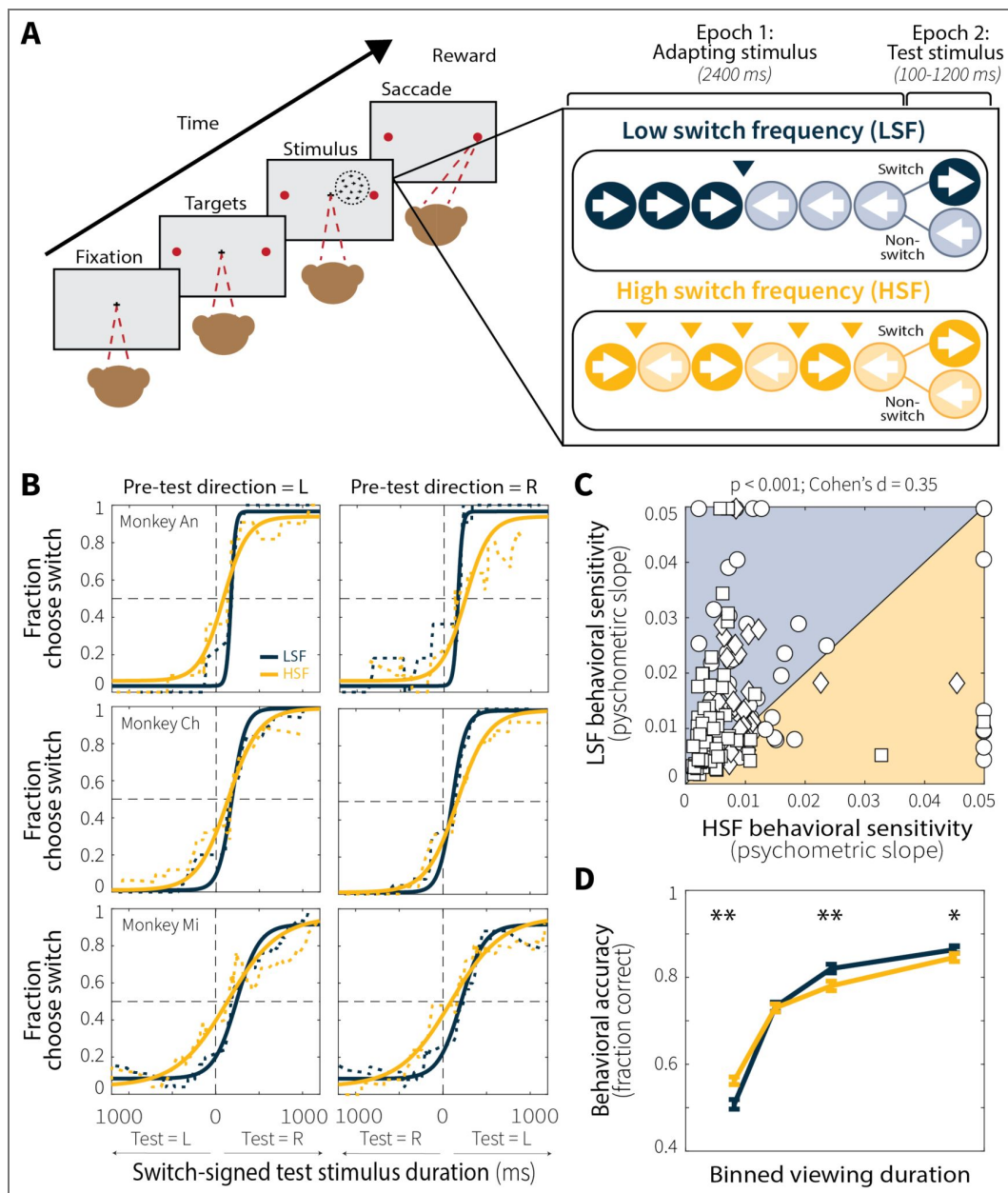


Fig. 2. Manipulating context stability in a random-dot motion task affects evidence-accumulation behavior.

(A) Each trial consisted of an adapting stimulus (2400 ms) followed immediately by a test stimulus (100–1200 ms, drawn from an exponential distribution). During the adapting stimulus, random-dot motion switched directions at either a low (LSF; 1 switch) or high (HSF; 5 switches) frequency, creating two context-stability conditions. Between epochs, there was a 50% probability of an additional change in motion direction, producing switch and non-switch trials. After presentation of the test stimulus, the fixation point was extinguished and the monkey reported the final direction of motion with a saccade to the corresponding choice target. (B) Running average (square window size=5 trials, sorted by test-stimulus duration) of behavioral choice data (dotted line) and psychometric fits (solid lines) for representative sessions from each monkey. LSF (blue) and HSF (orange) conditions were fit separately. Shallower slopes indicate decreased perceptual sensitivity as a function of viewing time. (C) Across-session comparison of fitted slopes for LSF and HSF conditions (Monkey An = 51 sessions, circles; Ch = 34, diamonds; Mi = 76, squares). p -value is from a Wilcoxon signed-rank test for equal medians. (D) Behavioral performance on switch trials as a function of test-stimulus duration, which was binned (100–225 ms, 225–375 ms, 375–600 ms, and 600–1200 ms) to include roughly equal numbers of trials (mean $\pm$ SEM across sessions). Asterisks denote bins for which LSF and HSF performance differed significantly (Wilcoxon signed-rank test for equal medians; * $p < 0.05$; ** $p < 0.001$).

Results

We trained three rhesus macaques (An, Ch, and Mi) to perform a version of a random-dot motion direction-discrimination task that combined reversals in dot-motion direction (Glaze et al., 2015; Fig. 2A) with elements of an adaptation-test paradigm (Van Wezel & Britten, 2002). Each trial consisted of two sequential epochs: an adapting epoch (Epoch 1) and a test epoch (Epoch 2). During the adapting epoch (2400 ms), the random-dot motion stimulus abruptly switched direction at either a low (LSF; 1 switch, occurring at 1200 ms) or high (HSF; 5 switches, occurring every 400 ms) frequency. The two adapting (“context-stability”) conditions were designed such that the total exposure time to each adapting motion direction was matched for low and high switch frequencies (1200 ms per direction). Immediately following the adapting epoch, the test stimulus was shown for a duration that was sampled independently on each trial from a truncated exponential distribution (100–1200 ms), requiring the monkeys to accumulate evidence over varied and unpredictable timescales. At the end of this duration, the test stimulus and fixation point were extinguished simultaneously, cueing the monkeys to report the final motion direction by making a saccade to the corresponding choice target.

Critically, for both low and high switch-frequency conditions there was a 50% probability that the motion stimulus changed direction between the adapting and test epochs. This design ensured that switch and non-switch trials were equally likely, and thus that the adapting stimulus provided no information about the correct test-stimulus direction. Moreover, the transition between adapting and test epochs was marked by predictable timing and a reduction in motion coherence (to ensure that the task was difficult enough to require sustained evidence accumulation), implying that the monkeys could, in principle, learn to ignore the adapting stimulus altogether. Instead, as detailed below, systematic effects of context stability on behavior, neural activity, and pupil modulations point to inherent ways in which the brain uses recent stimulus history to shape perception.

Context stability modulates evidence-accumulation behavior

To quantify if and how the monkeys’ decisions depended on recent temporal context, we fit their choice behavior using a time-dependent logistic function (Equation 1) that predicted the probability of reporting a switch in motion direction as a function of test-stimulus viewing time and trial type (switch or non-switch; Fig. 2B). This model captured behavior well, with Tjur’s pseudo- R^2 values that exceeded those computed by shuffling the association between test-stimulus durations and switch/non-switch trial types (Extended Data Fig. 1).

The adapting stimulus affected the monkeys’ decisions in several ways. On the shortest-duration switch trials, monkeys tended to make choices congruent with the final direction of the adapting stimulus, reflected in lower accuracy and a rightward shift of the psychometric functions (examples shown in Fig. 2B; mean $\pm$ SEM shift across sessions = 168.25 $\pm$ 8.38 ms for Monkey An, 110.79 $\pm$ 9.76 ms for Monkey Ch, 138.46 $\pm$ 10.35 ms for Monkey Mi). It was not until after ~140 ms that their choices began to more strongly reflect the direction of the test stimulus. Together, these effects indicate that the monkeys were not ignoring the adapting stimulus but instead were using at least some of its directional information to inform their decisions, particularly when given only brief exposure to the test stimulus. As the duration of the test stimulus increased, performance accuracy also tended to increase, including more “switch” choices on switch trials, suggesting that the monkeys’ decisions became more strongly dependent on the test as opposed to the adapting stimulus.

Critically, despite identical test-stimulus properties, durations, and equal probabilities of a behaviorally relevant switch, the rate of increase in accuracy as a function of viewing time differed between context-stability conditions. Psychometric slopes were shallower for the high relative to low switch-frequency condition (Fig. 2B–C; individual animals in Extended Data Fig. 2), reflecting diminished sensitivity to incoming evidence as a function of viewing time. As a consequence, performance on switch trials showed a “crossover” dynamic, such that accuracy was higher in the high relative to low switch-frequency condition at short durations but higher in the low relative to high switch-frequency condition at long durations (Fig. 2D; pairwise comparison

in [Extended Data Fig. 3A](#)). This pattern of behavior is consistent with adaptive adjustments in evidence accumulation, in which a higher leak rate or lower saturating non-linearity facilitates performance on short switch trials by downweighting pre-change evidence but limits performance on longer-duration switch trials as more recent, relevant evidence is also discounted. These context-dependent temporal dynamics are inconsistent with a simple shift in the timing of the onset of evidence accumulation (either advanced or delayed relative to test-stimulus onset), which would shift the psychometric curves but not change their slopes. Instead, these slope changes show that recent temporal context can shape the temporal dynamics of decision formation.

Context stability shapes sensory adaptation in MT

We recorded activity from 155 MT single units while the three monkeys performed the task (Monkey An = 55, Ch = 13, Mi = 87). As expected, given that stimulus properties were tailored to each neuron (see Methods), the activity of both individual units ([Fig. 3A](#)) and the recorded population ([Fig. 3B](#)) were robustly modulated by motion direction, showing strong responses to motion in each cell's preferred direction and weak responses in the opposite, anti-preferred ("null") direction that followed stimulus switching dynamics. Given the lack of responsiveness to null motion and the unbalanced exposure to the final motion direction on non-switch trials (i.e., an additional 800 ms at low relative to high switch frequency), unless otherwise noted the analyses of MT neural activity were restricted to preferred-motion switch trials, which were most informative for characterizing context-dependent evidence encoding.

Temporal-context stability systematically affected responses to preferred motion during the test epoch, with reduced activity following an adaptor that switched at a high versus low frequency ([Fig. 3C](#); individual animals in [Extended Data Fig. 4A](#)). Differences in neural activity between the two context-stability conditions were best characterized by a change in overall response magnitude rather than a change in temporal dynamics ([Extended Data Fig. 4B](#)). Moreover, these differences varied systematically with direction selectivity. MT single units with stronger direction selectivity, defined as larger differences between responses to preferred and null motion, showed greater differences in activity between switch-frequency conditions ([Fig. 3D](#)). In other words, neurons that encoded motion evidence more strongly and more selectively were also more sensitive to the temporal statistics and stability of that evidence, exhibiting larger context stability-dependent adjustments.

These context-dependent differences in MT neural activity did not reflect baseline differences in responsiveness that persisted throughout the low and high switch-frequency blocks. Rather, they emerged over the course of each trial, consistent with stimulus-specific sensory adaptation. Specifically, the average population responses to the first presentation of preferred motion during the adapting stimulus were initially matched between conditions ([Fig. 4A](#); although one animal showed an early difference, possibly reflecting learned expectation about switching dynamics within blocks, [Extended Data Fig. 5](#)). Context-dependent differences developed with repeated stimulus presentation within a trial, culminating in robust differences in neural responses to the test stimulus between low and high switch-frequency conditions ([Fig. 4A,B](#)).

The effect of context stability on evidence encoding was consistent across distinct patterns of adaptation. Whereas most MT single units showed persistent response decrements with repeated preferred-motion presentations ([Fig. 4C](#), *bottom*), some exhibited response increments ([Fig. 4C](#), *top*). We classified these two groups of neurons as "adapting" and "facilitating," respectively, based on changes in baseline-subtracted, normalized responses between the first presentation of preferred motion and preferred motion during the test stimulus (see Methods). Regardless of whether neurons were adapting or facilitating, both groups showed comparable magnitudes of response change as a function of preferred-motion stimulus presentation ([Fig. 4D](#)). Notably, the majority of this response change, particularly for adapting neurons, occurred between the first and second presentations of preferred motion, with relatively little additional change across subsequent presentations. This pattern suggests that adaptation dynamics are shaped by temporal features of the adapting stimulus, such as the frequency of direction switches, as opposed to the

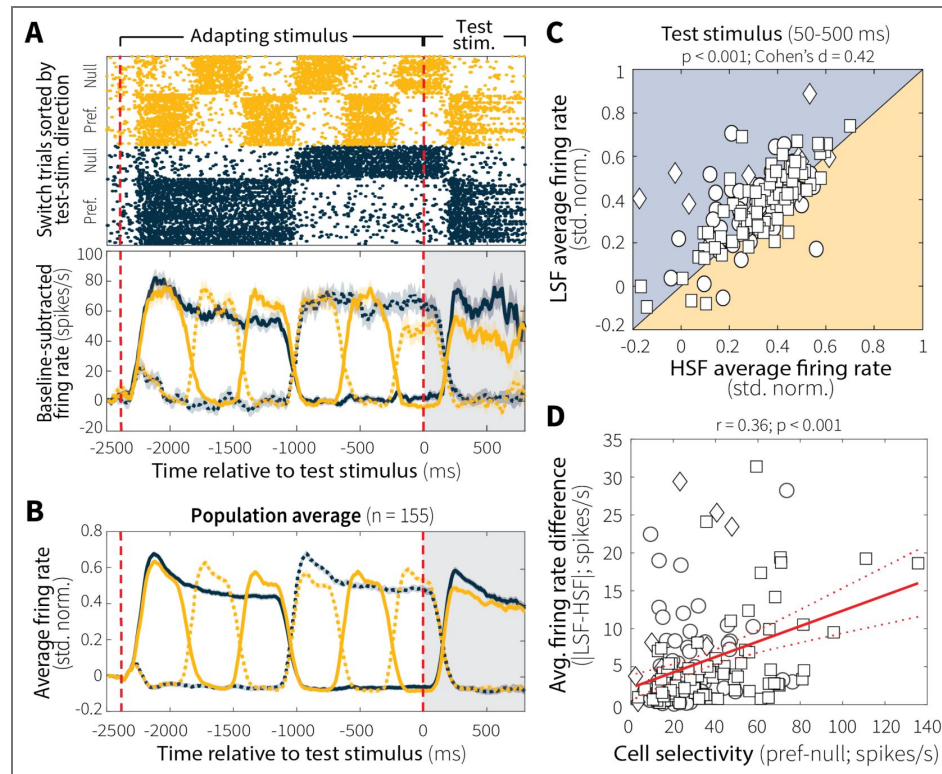


Fig. 3. Context stability modulates motion-evidence encoding in MT.

(A) Raster plot (top) and baseline-subtracted average firing rate (bottom; mean $\pm$ SEM) for switch trials from a representative MT single unit on low (LSF; blue) and high (HSF; orange) switch-frequency trials. Solid and dashed lines indicate responses to motion in the unit's preferred and null directions, respectively, defined relative to motion direction during the test stimulus. (B) Baseline-subtracted, normalized firing rate averaged across all recorded MT single units ($n = 155$; mean $\pm$ SEM). (C) Pairwise comparison of average single-unit activity during the test stimulus (50–500 ms) for LSF and HSF preferred-motion switch trials (Monkey An = 55 single units, circles; Ch = 13, diamonds; Mi = 87, squares). p -value is from a Wilcoxon signed-rank test for equal medians. (D) Pearson's correlation between MT direction selectivity and the magnitude of LSF–HSF response differences during the test stimulus (50–500 ms). Red solid and dashed lines indicate a linear fit $\pm$ 95% CI.

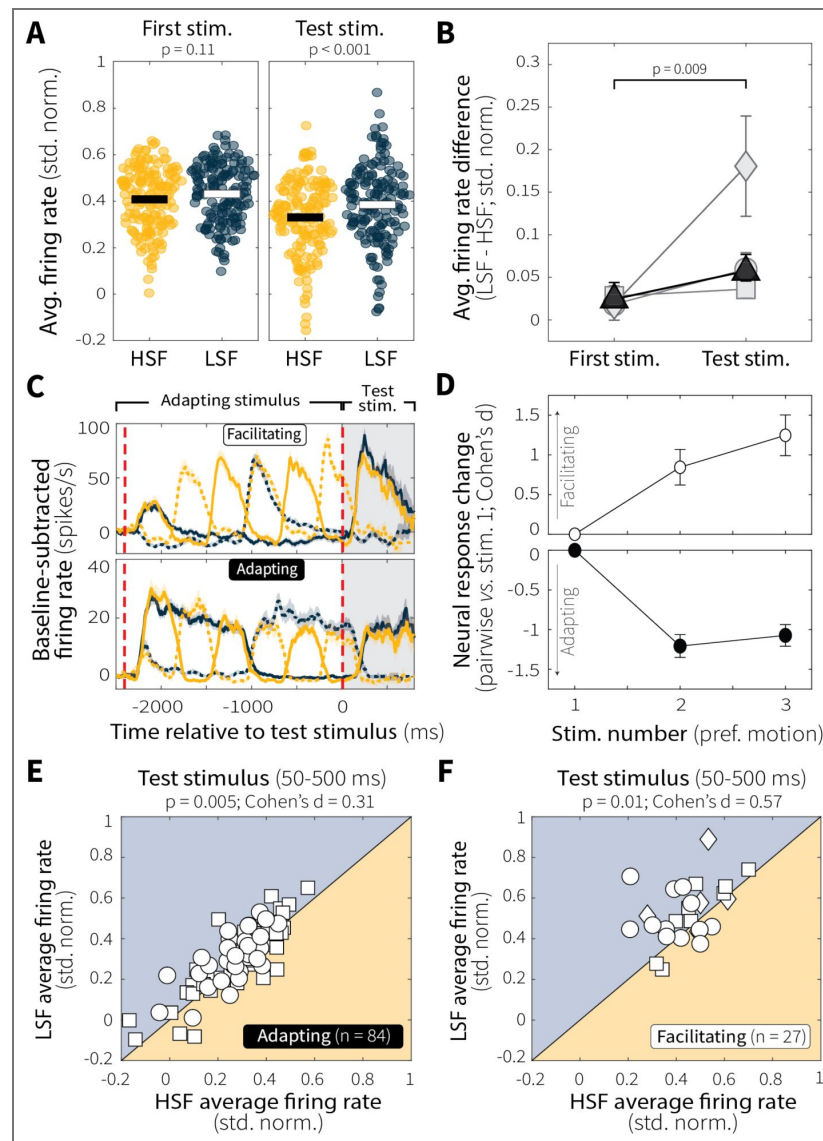


Fig. 4. Context-dependent differences in MT evidence encoding emerge following repeated preferred-motion stimulus presentation.

(A) Responses of MT single units (points) to the first (adapting-stimulus onset) and final (test) presentations of preferred-motion stimuli on switch trials. Horizontal bars represent means. p -values are for two-sample t -tests for equal means of the two distributions. (B) Change in response differences (LSF-HSF) from the first (adapting-stimulus onset) and final (test) presentations of preferred-motion stimuli on switch trials. Individual animals are in gray; group average is in black; all points are mean $\pm$ SEM across units. p -value is from a two-sample t -test comparing the full distribution of response differences (LSF-HSF) from the two time intervals for all monkeys. (C) Example switch-trials responses from representative facilitating (top) and adapting (bottom) single units. Colors and line styles as in Fig. 2A. (D) Comparison of average baseline-subtracted and normalized responses across successive presentations of preferred motion within the adapting epoch for HSF switch trials (panel C, orange solid). For adapting (black, $n = 87$) and facilitating (white, $n = 27$) groups, responses to each preferred-motion presentation were compared to the first (adapting-stimulus onset) preferred-motion response using Cohen's d to quantify changes in response magnitude as a function of stimulus number. Points denote mean $\pm$ SEM across units. (E, F) Pairwise comparison of average single-unit responses to the test stimulus (50-500 ms) on switch trials during LSF versus HSF conditions for adapting (E) and facilitating (F) units. Different symbols represent data from different monkeys, as in Fig. 2C. p -values are from a Wilcoxon signed-rank test for equal medians.

number of preferred-motion stimulus presentations. Moreover, regardless of whether MT neurons were adapting or facilitating, both groups exhibited robust context-stability effects with diminished activity at high versus low switch frequency during the test stimulus (Fig. 4E,F). Thus, context-specific differences in evidence encoding arose from stimulus-specific sensory adaptation that unfolded over repeated preferred-motion exposure and was expressed across MT neurons with different response profiles.

Context-dependent evidence encoding in MT relates to evidence-accumulation behavior

To assess the behavioral relevance of context-dependent differences in MT neural responses, we first quantified how well individual neurons discriminated between preferred and null motion at low versus high switch frequency using ROC area (Fig. 5A). Across the population, ROC area tended to be lower for high versus low switch frequency (Fig. 5B; individual animals in Extended Data Fig. 6A), indicating decreased evidence discriminability in MT when the test stimulus was preceded by a more unstable adapting context. We then tested whether, across sessions, these differences in ROC area predicted context-dependent behavioral differences. Because ROC analyses were restricted to switch trials (which, unlike non-switch trials, have balanced exposure to the final adapting motion direction), we also restricted the behavioral analyses to switch trials, using percent correct instead of psychometric slope (which was based on all trials). We focused on trials 375–600 ms in duration, because that time bin exhibited robust context-dependent differences in accuracy that remained reliably above chance, but below ceiling, for both conditions (Fig. 2D and Extended Data Fig. 3B). We found that the differences in MT evidence discriminability (LSF minus HSF ROC area, in the epoch 200–400 ms following test-stimulus onset that showed the largest context-dependent differences in neural activity) correlated with context-stability differences in behavioral performance (LSF minus HSF percent correct for trials ending 375–600 ms following test-stimulus onset, Fig. 5C), such that greater differences in discriminability between low and high switch-frequency conditions predicted larger differences in performance. This effect was significant across animals, but driven primarily by Monkey An (Extended Data Fig. 6B).

To further explore this relationship, we divided sessions based on evidence-accumulation behavior. For sessions in which the monkeys had shallower psychometric slopes at high relative to low switch frequency (Fig. 5D), MT neurons exhibited a corresponding decrease in firing rate and discriminability on switch trials (pairwise comparisons for firing rates in Extended Data Fig. 7). Conversely, for all other sessions (Fig. 5E), we did not identify any reliable context-stability differences in MT evidence encoding or discriminability (pairwise comparisons for firing rates in Extended Data Fig. 7). This relationship was specific to correct trials and absent for incorrect trials, which showed no modulation based on evidence-accumulation behavior (Extended Data Fig. 8). A similar, behavior-dependent pattern in discriminability was also evident on non-switch trials (Extended Data Fig. 9). However, these context-dependent differences were restricted to a smaller and earlier time window (50–200 ms), likely reflecting activity driven primarily by the adapting rather than the test stimulus. Together, these results suggest that adaptation-driven differences in MT evidence encoding contribute, at least in part, to context-dependent evidence-accumulation behavior.

Context-dependent evoked pupil modulations relate to evidence-accumulation behavior

To probe whether processes beyond local sensory adaptation contribute to adaptive evidence accumulation, we examined both baseline and evoked pupil diameter as proxies for arousal-related modulations that have been shown to encode key features of adaptive perceptual decision making (Cheadle et al., 2014; Krishnamurthy et al., 2017; Murphy et al., 2021). Trial-wise baseline pupil diameter was correlated with the time required for monkeys to obtain fixation (Extended Data Fig. 10A), a relationship that was robust across sessions and animals (Extended Data Fig. 10B). However, baseline pupil diameter did not differ systematically between low and

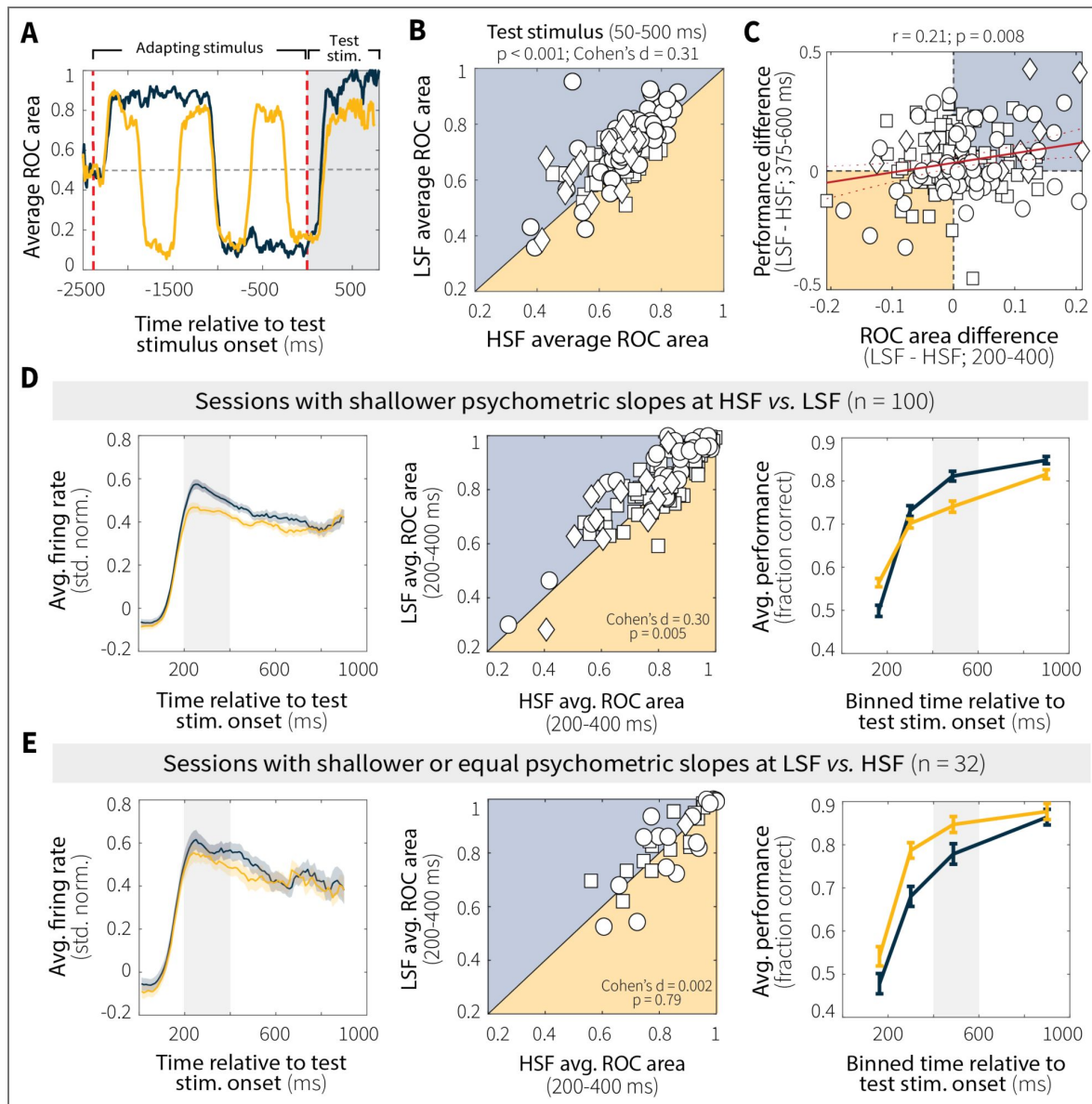


Fig 5. MT neural activity relates to evidence-accumulation behavior.

(A) Average ROC area over time for a representative MT single unit showing discriminability between preferred and null motion during low (LSF, blue) and high (HSF, orange) switch-frequency switch trials. (B) Pairwise comparison of average ROC area during the test stimulus (50–500 ms) for LSF and HSF switch trials. p -value is from a Wilcoxon signed-rank test for equal medians. (C) Context stability-dependent differences in behavioral performance differences (LSF–HSF, measured as the difference in accuracy for trials that ended 375–600 ms after test-stimulus onset; gray band in D and E, left panels) plotted as a function of MT single-unit discriminability (LSF–HSF ROC area, 200–400 ms after test-stimulus onset; gray band in D and E, right panels). Red solid and dashed lines indicate a linear fit $\pm 95\%$ CI. p -value is for the Pearson's correlation coefficient ($H_0: r = 0$). (D) Sessions with shallower psychometric slopes at HSF relative to LSF. Columns show (left) MT evidence encoding (mean $\pm$ SEM baseline-subtracted, normalized firing rates averaged across units), (middle) MT evidence discriminability (average ROC area, 200–400 ms after test-stimulus onset; p -values from a Wilcoxon signed-rank test for equal medians), and (right) behavioral performance in stimulus-duration bins with roughly equal numbers of switch trials (mean $\pm$ SEM across sessions). (E) Same as D, but for sessions with shallower or equal slopes at LSF relative to HSF. Different symbols in the four scatterplots represent data from different monkeys, as in Fig. 2C.

high switch-frequency conditions (Extended Data Fig. 10C [↗](#)), suggesting that it captured global as opposed to context-dependent fluctuations in arousal or task engagement. We therefore focused on task-evoked, as opposed to baseline, pupil responses in subsequent analyses.

Evoked pupil responses were modulated by context stability. Following a brief, stereotyped constriction caused by the abrupt luminance increase at motion-stimulus onset, differences in pupil diameter emerged between low and high switch-frequency conditions (Fig. 6A [↗](#)). To quantify these context-dependent effects, we used sliding-window linear regressions (Equation 2 [↗](#)) to estimate differences in evoked pupil diameter (LSF–HSF), while controlling for baseline pupil diameter (because evoked and baseline diameter tend to be inversely related; Mathôt et al., 2018 [↗](#)). For both monkeys (Monkey Ch was excluded from pupil-related analyses; see Methods), the effect of context-stability on evoked pupil diameter increased over the course of the adapting stimulus (Fig. 6B [↗](#); individual animals in Extended Data Fig. 11A [↗](#)). This result is consistent with the idea that evoked pupil responses track context-dependent expectations or anticipatory processes associated with behaviorally relevant stimuli.

Context-dependent differences in evoked pupil diameter were relevant to behavior. Specifically, differences in evoked pupil diameter (LSF–HSF) showed correlations with behavioral sensitivity (LSF–HSF psychometric slope) that varied over the course of the trial and peaked just before the onset of the test stimulus (Fig. 6C [↗](#)). Around this peak (~500–0 ms before test-stimulus onset), average context-dependent differences in evoked pupil diameter reliably predicted subsequent differences in behavioral sensitivity across sessions (Fig. 6D [↗](#); individual animals in Extended Data Fig. 11B [↗](#)). As such, sessions with larger differences in evoked pupil diameter also tended to show larger context-dependent differences in behavioral sensitivity.

Adaptation and arousal-related mechanisms are jointly and differentially recruited across sessions

To directly compare the relative contributions of MT sensory adaptation and arousal-related mechanisms to adaptive evidence-accumulation behavior, we separately incorporated trial-wise measures of MT neural activity and evoked pupil diameter into the psychometric functions (Equation 3 [↗](#)). This approach allowed us to quantify how strongly each signal modulated behavioral sensitivity as a function of context stability (LSF–HSF) on a session-by-session basis.

Both neural and pupil terms improved model fits, as evidenced by increased explanatory power relative to the behavior-only model in both animals (Fig. 7A [↗](#)). However, the relative improvement attributable to neural activity versus evoked pupil diameter differed between monkeys, with Monkey An showing a larger difference between neural versus pupil terms than Monkey Mi (Wilcoxon rank-sum test on neural–pupil differences in Tjur's pseudo- R^2 ; $p < 0.001$, Fig. 7A [↗](#)), implying joint, but flexible recruitment of adaptation and arousal-related mechanisms.

This flexible recruitment was evident across sessions, where the influence of context-dependent MT neural activity and evoked pupil diameter on behavioral sensitivity varied considerably, with neither term reliably predominating over the other (Fig. 7B [↗](#)). Session-wise, context-dependent contributions from neural and pupil terms were uncorrelated with each other, consistent with the two signals contributing separately rather than redundantly or in opposition (Fig. 7B [↗](#)). In further support of this idea, context-dependent modulations of pupil size were not related systematically to context-dependent modulations of MT neural activity or discriminability (Extended Data Fig. 11C [↗](#), D). Together, these findings suggest that sensory adaptation and pupil-linked arousal both contribute to adaptive evidence accumulation, likely operating at different processing stages, with their relative weighting varying substantially across sessions and animals.

Discussion

Flexible evidence accumulation is critical for effective decision making in dynamic environments. Many algorithmic accounts propose that this flexibility arises from context-dependent adjustments to the process of evidence accumulation itself. However, the brain contains many adaptive mechanisms at different stages of the decision process that could contribute to this behavioral

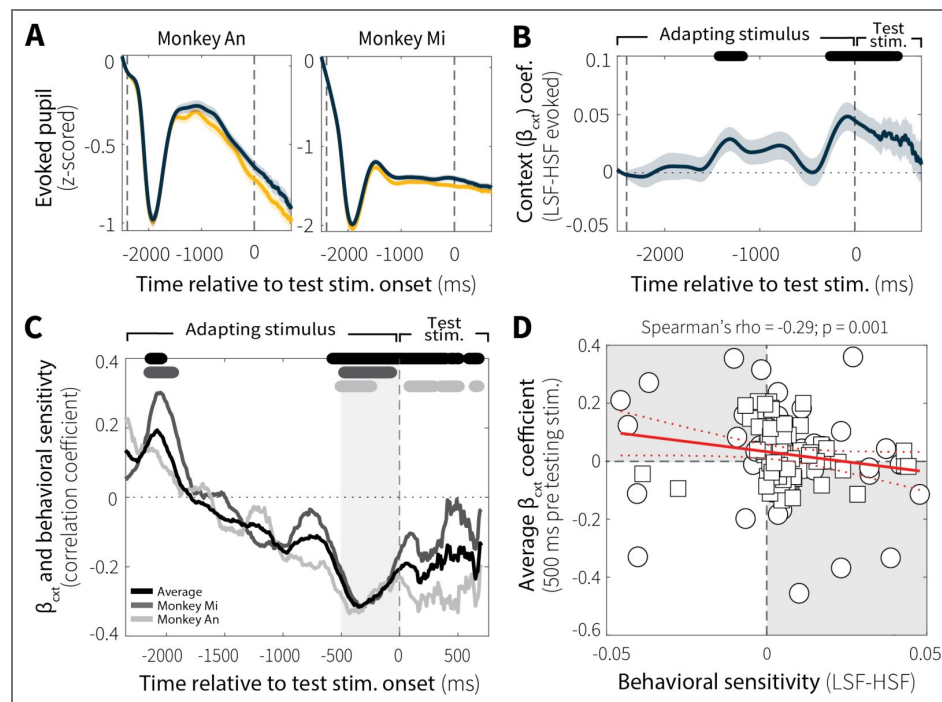


Fig. 6. Evoked pupil responses depend on context stability and relate to evidence-accumulation behavior.

(A) Average evoked pupil responses over time (z-scored per session) for low (LSF; blue) and high (HSF; orange) switch-frequency conditions (mean $\pm$ SEM across all sessions), shown separately for the two monkeys. (B) Sliding-window regression coefficients (β_{ctx} ; mean $\pm$ SEM across switch and non-switch trials for all sessions from both monkeys) estimating the effect of context-stability condition (LSF-HSF) on pupil diameter during stimulus viewing while controlling for baseline pupil diameter. Black bar (top) indicates windows in which β_{ctx} differed significantly from zero ($p < 0.05$, uncorrected for multiple comparisons). (C) Spearman's rank correlation coefficient between behavioral sensitivity (LSF-HSF psychometric slope) and β_{ctx} coefficients from individual sessions as a function of time relative to test-stimulus onset (computed in 100 ms bins with 10 ms steps). Data from individual animals are shown separately (Monkey An: light grey; Monkey Mi: dark grey) along with the across-animal average (black). Corresponding grayscale bars at the top indicate time points with significant correlations ($p < 0.05$, uncorrected for multiple comparisons). (D) Spearman's rank correlation between average evoked pupil diameter differences (β_{ctx} , -500-0 ms relative to test-stimulus onset; gray band in C) and behavioral sensitivity differences (LSF-HSF slope) across sessions. Red solid and dashed lines indicate a linear fit $\pm$ 95% CI. p -value is for the Spearman's correlation coefficient ($H_0: \rho = 0$).

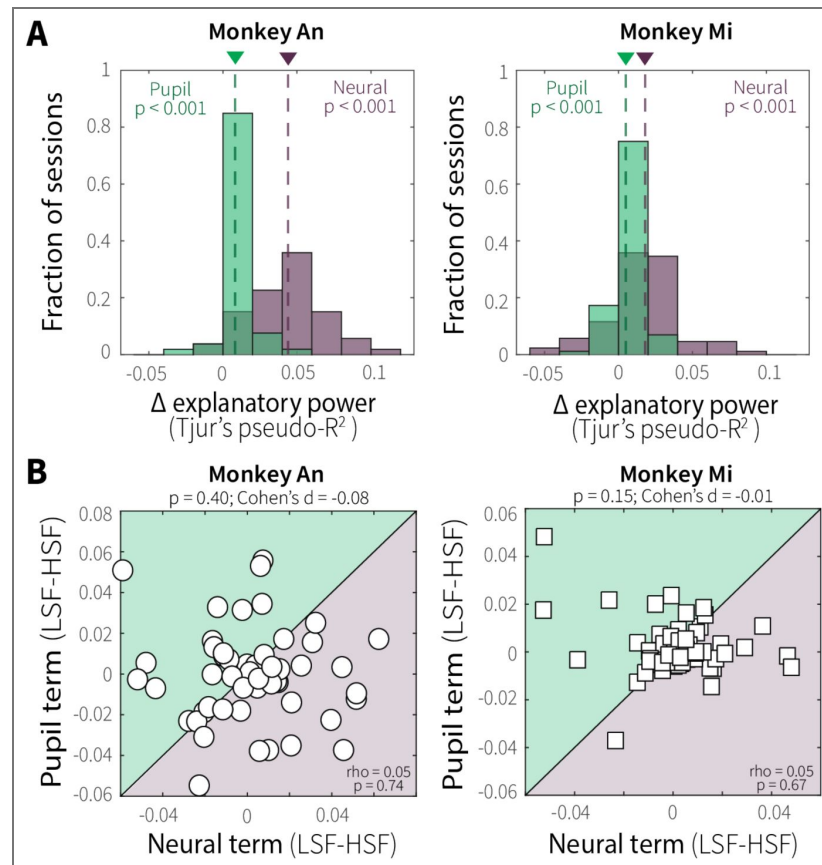


Fig. 7. Context-stability jointly and differentially recruits adaptation and arousal-related mechanisms across sessions.

(A) Distributions of differences in explanatory power (Tjur's pseudo- R^2) for models fit with versus without the neural (purple) or pupil (green) term across sessions for Monkey An (left) and Monkey Mi (right). Mean values are indicated by dashed lines and corresponding colored triangles. p -values are from a one-sample sign test for H_0 : fit parameters came from a distribution with a median of zero. (B) Context-stability differences (LSF-HSF) in neural (MT firing rate; abscissa) and pupil (evoked diameter; ordinate) model terms for Monkey An (left) and Monkey Mi (right). Points are fits to data from individual sessions/units from logistic models that quantified how much the slope of time-dependent psychometric functions covaried with the given neural or pupil term, computed separately for LSF and HSF trials. Shaded quadrants denote relative dominance of neural (purple) or pupil (green) effects. p -values are from a Wilcoxon signed-rank test for equal medians (top) and Spearman's correlation (bottom).

flexibility. Motivated by this idea, we manipulated the temporal stability of recent sensory experience and simultaneously measured both neural activity in the middle temporal area (MT) and pupil-linked arousal. We identified two complementary mechanisms that could support flexible evidence accumulation, one involving sensory adaptation and the other involving arousal-related neuromodulation (Fig. 1). Our results suggest that these mechanisms operate separately, likely at different processing stages, providing a new view of the distributed neural substrates for flexible evidence accumulation in dynamic environments.

Our MT findings indicate that context-dependent changes in how choice accuracy evolves as a function of viewing time, which is a behavioral signature of flexible evidence accumulation, can arise, at least in part, from context-dependent differences in sensory encoding. Both motion-evidence encoding and discriminability of MT responses during the test-stimulus tended to be reduced following exposure to more temporally unstable adapting stimuli (Fig. 3A-C). These reductions were not sustained differences throughout the two context-stability conditions but instead emerged on each trial through repeated presentations of preferred motion at different switch frequencies, consistent with stimulus-specific sensory adaptation (Fig. 4A-B). This stimulus-specific adaptation varied systematically across individual units, such that more direction-selective cells were more affected by differences in stimulus temporal statistics, which is consistent with previous findings that more active MT neurons adapt more strongly (Fig. 3D; Van Wezel & Britten, 2002).

It is well established that sensory adaptation can depend on stimulus statistics such as the mean, variance, and salience of sensory inputs (Dean et al., 2005; Fairhall et al., 2001) and even changes in those statistics (Kastner & Baccus, 2011; Tikidji-Hamburyan et al., 2015). Here, we extend these findings by demonstrating that adaptation can also depend on the rate of change of the stimulus statistics. This kind of temporal dynamic-dependent adaptation has been observed in retinal ganglion cells (Wark et al., 2009). Our work extends those findings to non-human primate cortical neurons with causal links to behavior (Hanks et al., 2006; Newsome & Pare, 1988), in particular showing that context stability-dependent adaptation in MT shapes the sensory evidence available for time-evolving perceptual decisions.

We also observed heterogeneity in how MT neural activity changed over the course of the adapting stimulus, with some cells exhibiting progressive response attenuation and others facilitation (Fig. 4C-D). A plausible mechanism for this facilitation is that adaptation alters the local excitation-inhibition balance (Higley & Contreras, 2007; Solomon & Kohn, 2014). In particular, repeated presentation of null motion may weaken opponent input, producing disinhibition of preferred-motion responses (Kohn & Movshon, 2003). Consistent with this account, facilitating neurons showed increased responses to preferred motion following null-motion adaptation (Extended Data Fig. 12). This diversity of adaptation profiles mirrors observations in retinal ganglion cells (Kastner & Baccus, 2011, 2013), primary auditory cortex (Seay et al., 2020), and somatosensory cortex (Adibi et al., 2013; Cohen-Kashi Malina et al., 2013; Derdikman et al., 2006; Kheradpezhough et al., 2017), and recent work suggests that whether a neuron adapts or facilitates may itself be a flexible property that depends on recent sensory experience (Dobler et al., 2024). Nevertheless, regardless of the adaptation profile, context stability shaped MT activity and discriminability in ways that corresponded to context-dependent differences in evidence-accumulation behavior across sessions.

These adaptation-driven changes in MT evidence encoding were relevant to behavior but not fully prescriptive of context-dependent adjustments in evidence accumulation. For instance, for sessions in which the monkeys showed behavioral evidence of context-dependent evidence-accumulation (i.e., reduced sensitivity as a function of viewing time at high relative to low switch frequency; Fig. 5D), there was a corresponding change in MT evidence encoding (e.g., reduced neural activity at high relative to low switch frequency; Fig. 5D). However, sessions showing the opposite pattern of behavior did not show opposite neural effects (Fig. 5E). Thus, while sensory adaptation supports flexible evidence accumulation by shaping the fidelity of sensory input, additional mechanisms likely contribute to how that information is dynamically weighed and accumulated over time.

Consistent with this idea, fluctuations in task-evoked pupil size, an index of arousal-related neuromodulation, were also related systematically to temporal-context stability. Differences in evoked pupil diameter emerged between low and high switch-frequency conditions over the course of the adapting epoch, peaking just before test-stimulus onset (Fig. 6B). This time course suggests that evoked pupil dynamics may be tracking expectations about environmental statistics to prepare downstream decision circuits for the behaviorally relevant stimulus, consistent with previous work (Filipowicz et al., 2020). Across sessions, the magnitude of context-dependent differences in evoked pupil diameter predicted corresponding differences in behavioral sensitivity (Fig. 6C-D). Specifically, larger pupil diameter corresponded to less-efficient evidence accumulation, a result consistent with previous work relating larger pupil diameter to increased variability (Murphy et al., 2014) and individual differences in the temporal dynamics of evidence accumulation (Keung et al., 2019). These findings join a growing body of work showing that pupil-linked arousal contributes to adaptive mechanisms that shape behavior to match current environmental demands.

Our findings further suggest that these pupil-linked arousal mechanisms operate separately from sensory adaptation to support flexible evidence accumulation. Adding trial-wise MT neural activity or evoked pupil diameter each increased explanatory power relative to behavior-only models, demonstrating that both signals provided behaviorally relevant information (Fig. 7). However, the fact that their relative contributions were uncorrelated across sessions and that evoked pupil responses were unrelated to MT activity suggests that sensory adaptation and pupil-linked arousal may shape behavior through distinct mechanisms at different processing stages. That said, given the general unreliability of trial-wise data, particularly pupil responses, we cannot rule out the possibility that the lack of correlation between adaptation and arousal-related contributions results, in part, from this inherent variability as opposed to true independence of the underlying mechanisms.

One possible interpretation of these arousal-related signals is that they are related to neuromodulatory mechanisms that control the efficiency or leakiness of downstream accumulator circuits, such as those in the lateral intraparietal area (LIP), frontal eye field (FEF), or superior colliculus (SC), that read out MT activity. Alternatively, or in addition, both MT adaptation and pupil-linked arousal could be independently regulated by prefrontal regions, such as the dorsolateral prefrontal cortex (dlPFC), anterior cingulate cortex (ACC), or orbitofrontal cortex (OFC), that monitor task performance and environmental state (Behrens et al., 2007; McGuire et al., 2014; O'Reilly et al., 2013) and project both to brainstem arousal systems (Breton-Provencher & Sur, 2019; Schwarz & Luo, 2015) and to sensory cortex (Banerjee et al., 2020; Gilbert & Li, 2013; Murphy et al., 2021). Either possibility supports a framework in which sensory adaptation provides a “bottom-up” mechanism that modulates evidence encoding based on recent stimulus statistics, whereas pupil-linked arousal provides a “top-down” signal that adjusts how that encoded evidence is accumulated over time according to expectations about environmental stability. Together, these complementary processes may jointly implement the context-dependent nonlinearities described by normative models of evidence accumulation in dynamic environments (Glaze et al., 2015; Murphy et al., 2021; Fig. 1), with variability in their relative contributions across sessions and animals suggesting flexible recruitment that may depend on factors such as training, experience, or strategy.

Our conclusions are tempered by several limitations. First, we do not know the exact algorithmic nature of the context-dependent effects we measured on evidence accumulation behavior (i.e., differences in how accuracy evolves as a function of viewing time), which could involve a leak, saturating non-linearity, and/or other form of adjustment. Disentangling these alternatives necessitates additional behavioral testing and analyses (e.g., reverse-correlation analyses, which would require interspersing lower-coherence test stimuli than were used in this study) under a broader range of conditions. Second, we do not know where the context-dependent adaptation measured in MT originates. In principle, these effects could, at least in part, be inherited from upstream visual areas, including areas V1 (Born & Bradley, 2005; Churchland et al., 2005; Movshon & Newsome, 1996; Rodman et al., 1989) and V2 (DeYoe & Van Essen, 1985), where

adaptation to motion is also known to occur. Third, our recordings sampled single units across sessions and thus did not capture population-level changes in MT activity. Adaptation can influence activity beyond single neurons, including changes in correlated variability within neural populations that can affect decision making (Solomon & Kohn, 2014 [↗](#); Whitmire & Stanley, 2016 [↗](#)). Moreover, correlated variability in cortex can covary with pupil size and LC-NE activity (Joshi & Gold, 2022 [↗](#); Vinck et al., 2015 [↗](#)), providing another potential link between arousal and sensory encoding that merits further study. Fourth, downstream cortical (Ding & Gold, 2012 [↗](#); Kim & Shadlen, 1999 [↗](#); Shadlen & Newsome, 2001 [↗](#)) and subcortical (Ding & Gold, 2013 [↗](#); Horwitz & Newsome, 1999 [↗](#); Lovejoy & Krauzlis, 2010 [↗](#)) areas that read out MT activity to form the decision may also contribute to flexible adjustments to the process of evidence accumulation. Future work combining further behavioral testing and simultaneous recordings and/or causal manipulations across sensory, decision, and arousal-related regions will be essential to disentangle how multiple distributed mechanisms implement adaptive evidence accumulation to support flexible decision making.

In summary, this work contributes to a broader reconceptualization of classic, hierarchical views of sensorimotor pathways that cast primary sensory cortex as a static encoder supplying input to flexible downstream decision processes that guide behavior. Instead, a growing body of evidence suggests that these early sensory areas can provide adaptive, context-dependent signals shaped by recent experience and environmental structure that can impact behavior (Waiblinger et al., 2022 [↗](#), 2025 [↗](#)). Our demonstration of separable, behaviorally relevant contributions from sensory adaptation in MT and pupil-linked arousal supports this view, indicating that flexible decision making in dynamic environments emerges from distributed mechanisms that include intrinsically adaptive signals in early sensory cortex that complement adjustments in downstream processes.

Materials and Methods

Three adult male rhesus monkeys (*Macaca mulatta*) participated in this study. Two of the monkeys (An and Ch) had extensive prior training on a standard (i.e., non-switching) version of the random-dot motion direction-discrimination task. All training, surgery, and experimental procedures conformed to the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the University of Pennsylvania Institutional Animal Care and Use Committee (protocol #806027).

Behavioral task

The behavioral task (Fig. 1A [↗](#)) combined a change-point variant of the random-dot motion task (Glaze et al., 2015 [↗](#)) with adaptation-test elements from traditional MT sensory adaptation experiments (Van Wezel & Britten, 2002 [↗](#)). Briefly, each trial consisted of two sequential epochs: an adapting-stimulus epoch (2400 ms) and a test-stimulus epoch (100–1200 ms, sampled independently from an exponential distribution with a mean of 500 ms). During the adapting epoch, a continuous random-dot motion stimulus was presented, within which motion direction switched at either a low (LSF; 1 switch) or high (HSF; 5 switches) frequency, creating two context-stability conditions. Change-point frequencies were designed so that total exposure time to each motion direction was matched across conditions (1200 ms per direction). Between epochs, there was a 50% probability of an additional change in motion direction, yielding an approximately equal number of switch and non-switch trials within each session. During the test epoch, the monkey maintained fixation and accumulated motion evidence over variable durations before reporting the final motion direction with a saccade to one of two choice targets. A juice reward was provided following correct choices.

For Monkey An (all 51 sessions), Mi (all 76 sessions), and Ch (11 out of 35 sessions), low (LSF) and high (HSF) switch-frequency conditions were run in blocks with counterbalanced order across sessions. The remaining sessions for Monkey Ch ($n = 24$) included interleaved LSF and HSF trials. Initial motion direction was counterbalanced across trials within each session. The adapting

stimulus used high motion coherence (70%) to drive robust MT neural responses and adaptation, whereas the test stimulus used lower motion coherence (50–60%, adjusted across sessions to maintain overall task performance near ~70–80% correct) to promote temporal integration.

Eye position was monitored with a video-based system (EyeLink, SR Research) sampled at 1000 Hz and used to enforce fixation during motion viewing, register the saccadic response, and provide reward/error feedback online based on comparisons between the saccade endpoint and target locations.

Analysis of behavioral data

Behavioral choice data were fit with a logistic function that modeled the probability of reporting a switch in motion direction as a function of test-stimulus duration and trial type (switch versus non-switch):

$$p(x) = \lambda + \frac{1 - 2\lambda}{1 + e^{-\beta_{\text{slope}} (x_1 - (\beta_{\text{stay-switch}} x_2 + \beta_{\text{right-left}} x_3))}} \quad (1)$$

where λ is the error (“lapse”) rate independent of motion information (which here could arise for numerous reasons, including not just lapses of attention but also possibly inappropriate associations with motion directions observed during the adapting epoch); β_{slope} governs the steepness of the psychometric function as a function of test-stimulus viewing time (x_1), signed by trial type (i.e., multiplied by 1 for switch trials and -1 for non-switch trials); $\beta_{\text{stay-switch}}$ captures biases related to reporting switch versus non-switch (x_2 ; switch = 1, stay = 0); and $\beta_{\text{right-left}}$ captures biases related to right versus left (x_3 ; right = 1, left = 0) choices. Here, “right” and “left” refer to all motion directions on the right or left side of vertical, respectively (e.g., motion directions between 270°–90° labeled “right,” and 90°–270° labeled “left”). Bias terms were encoded with respect to the x-intercept because the steep slopes can make the y-intercept difficult to estimate. We imposed an upper bound of 0.05 on β_{slope} , which captured the majority of variability across sessions. This formulation allowed for the construction of a time-dependent psychometric function in which β_{slope} reflects evidence accumulation over time. Low and high switch-frequency conditions were fit separately to allow for comparison of fitted slope parameters.

Electrophysiology

General surgical procedure and data acquisition methods have been described previous in detail (Law & Gold, 2008 [DOI](#)). Briefly, monkeys were prepared for experiments via surgical implantation of a head-holding device and recording cylinder. Area MT was targeted using stereotaxic coordinates, magnetic resonance imaging, and 3D reconstruction with BrainSight (Rogue Research, Inc). We searched for MT neurons with consistent spatial, direction, and speed tuning using a 99% coherence random-dot motion stimulus.

For each recording, multiple properties of the random-dot motion stimulus including aperture size and location, speed, and net motion direction were adjusted to match the tuning properties of the isolated single-unit and thus maximize stimulus-specific sensory adaptation. As such, exact stimulus parameters varied across sessions with stimulus sizes ranging from 4.6–15 degrees, locations within the contralateral visual hemifield ± 5 degrees from the point of fixation, speeds 2–8 degrees/s, and motion-directions distributed roughly evenly within circular space. The adapting stimulus began with preferred or anti-preferred motion chosen at random with equal probability and counterbalanced across trials.

Neural activity was recorded using polyamide-coated tungsten electrodes (FHC, Inc) or glass-coated tungsten electrodes (Alpha-Omega). Spike waveforms were stored and sorted offline using software from Plexon, Inc. In some sessions, multiple MT single units were recorded simultaneously ($n = 40$ units in 19 sessions).

Analysis of MT neural data

Firing rates were computed for each neuron and trial condition using a 100 ms sliding window advanced in 10 ms steps and aligned to the onset of the test stimulus. For each MT unit, firing rates were averaged across trial conditions, baseline subtracted on a trial-by-trial basis using responses from the 100 ms window immediately preceding adapting-stimulus onset as baseline, and normalized by the maximum firing rate observed during the 2400 ms adapting stimulus. For analyses requiring an average firing rate during the test stimulus, MT neural activity was averaged from a specified onset to either test-stimulus offset or the end of the specified time window, whichever came first. Only correct trials were included in analyses unless otherwise indicated.

Direction selectivity was quantified as the average difference between responses to preferred and anti-preferred (null) motion from 100 to 500 ms after adapting stimulus onset, corresponding to the overlapping stimulus period (400 ms) across switch-frequency conditions with a 100 ms delay.

To classify units as “adapting” versus “facilitating,” average baseline-subtracted and normalized neural activity was compared between the first presentation of preferred motion (200–400 ms following adapting stimulus onset) and preferred motion during the test stimulus (50–500 ms). Units were classified as adapting if responses decreased by >2.5% and as facilitating if responses increased by >2.5%, under both LSF and HSF conditions. We chose a liberal criterion for classification because we were interested in maintaining enough neurons in each group for statistical comparison of sensory adaptation as a function of preferred-motion stimulus presentation. Alternatively, using linear fits to peak preferred-motion responses and classifying neurons based on whether the fitted slope differed significantly from zero ($p < 0.05$), yielded 34 units (21.9%) with negative slopes classified as “adapting” and 4 units (0.3%) with positive slopes classified as “facilitating,” which are comparable to prior work in sensory cortical areas (Dobler et al., 2024 [DOI](#); Kheradpezhrouh et al., 2017 [DOI](#); Seay et al., 2020 [DOI](#)).

Analysis of pupil data

For each session, pupil data were aligned to the onset of the adapting stimulus and preprocessed by: (1) linear interpolation of missing values, (2) applying a first-order 5 Hz low-pass Butterworth filter, (3) removing and linearly interpolating samples exceeding ± 2 standard deviations from the mean, (4) subtracting the session-averaged response time course, and (5) z-scoring. For evoked-pupil analyses, only correct trials were included. For baseline pupil analyses, all trials were included so as not to disrupt the session’s temporal order.

Monkey Ch was excluded from pupil analysis because many sessions used interleaved (versus blocked) context-stability conditions, which removed the ability to establish trial-by-trial expectations of stability. Two sessions for Monkey An were also excluded because the pupil data were not properly saved.

To quantify context-stability effects on evoked-pupil dynamics, we applied sliding-window linear regressions (100 ms window, 10 ms slide; Equation 2 [DOI](#)) to estimate differences in evoked pupil response between LSF and HSF trials (β_2) while controlling for baseline pupil (β_1).

$$y = \beta_0 + \beta_{\text{baseline}} x_1 + \beta_{\text{cxt}} x_2 \quad (2)$$

where, β_{baseline} estimates influence of baseline pupil diameter (x_1 : 100 ms window preceding motion stimulus onset) and β_{cxt} indexes context-stability condition (x_2 : 1 = LSF, 0 = HSF).

Quantification of neural and pupil contributions to behavior

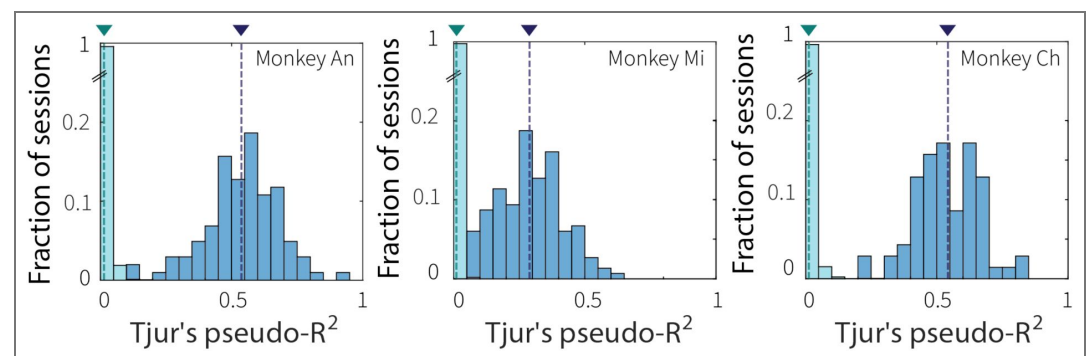
To quantify how neural and pupil measures contributed to behavioral sensitivity on a session-by-session basis, we added a term, $\beta_{\text{trial-wise}}$, to our time-dependent logistic function (Equation 3 [DOI](#)).

$$p(x) = \lambda + \frac{1 - 2\lambda}{1 + e^{-\beta_{\text{slope}} \beta_{\text{trial-wise}} (x_1 * x_4 - (\beta_{\text{stay-switch}} \lambda_2 + \beta_{\text{right-left}} \lambda_3))}} \quad (3)$$

where, $\beta_{\text{trial-wise}}$ estimates the contribution of trial-wise MT activity for neural fits or trial-wise evoked pupil diameter for pupil fits (x_4).

For these fits, MT activity was computed as the average firing rate during the test stimulus on each trial with a 50 ms delay. Because MT neurons exhibit minimal response to null motion, only trials ending in preferred motion were used for neural fits. Evoked pupil diameter was calculated as the average of the residuals (500 ms window preceding test-stimulus onset) from a linear fit with average pupil diameter and baseline pupil diameter. Otherwise, parameterization of the fits was as described in Equation 1. Given the lack of evidence for an interaction between pupil-linked arousal and sensory adaptation, neural and pupil terms were assessed as the sole contributor in separate fits. When assessing explanatory power, for each session, the reported pseudo- R^2 value was the average across pseudo- R^2 values from separate fits to low (LSF) and high (HSF) switch-frequency trials. Tjur's pseudo- R^2 is a coefficient of discrimination. As such, its value increases only when additional model terms improve the separation of predicted probabilities across outcomes, providing a conservative basis for comparing models of differing complexity.

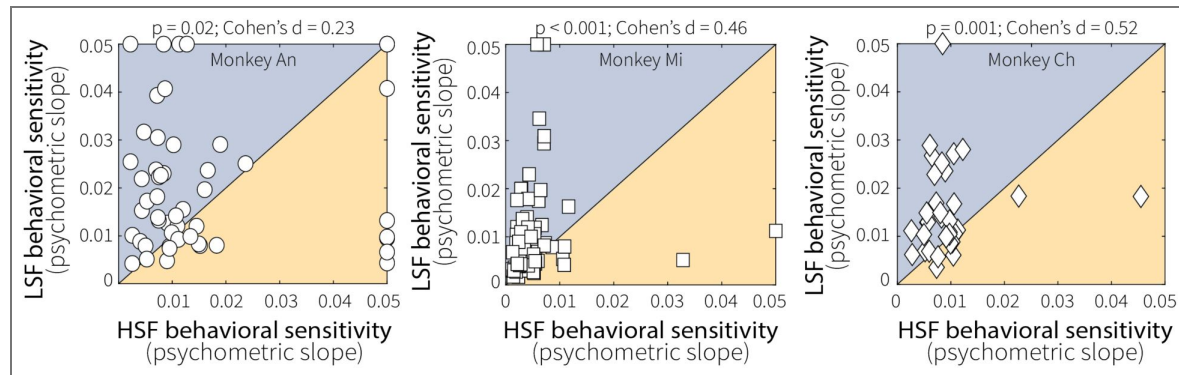
Supplementary figures



Extended Data Fig. 1. Time-dependent logistic model captures choice behavior. Distribution of Tjur's pseudo- R^2 values from empirical logistic fits (blue) compared with values from shuffled control ("null") fits (cyan) for each monkey, as indicated. Null fits were obtained by shuffling the association between test-stimulus durations and switch/non-switch trial types across 100 iterations per session. For each session, the reported pseudo- R^2 value was the average from separate fits to low and high switch-frequency trials. Dashed lines and triangles indicate mean values for each distribution. Enhanced explanatory power of the time-dependent logistic model versus the shuffled null was consistent across the three monkeys (Wilcoxon rank-sum test for equal medians: Monkey An: $p < 0.001$; Monkey Mi: $p < 0.001$; Monkey Ch: $p < 0.001$).

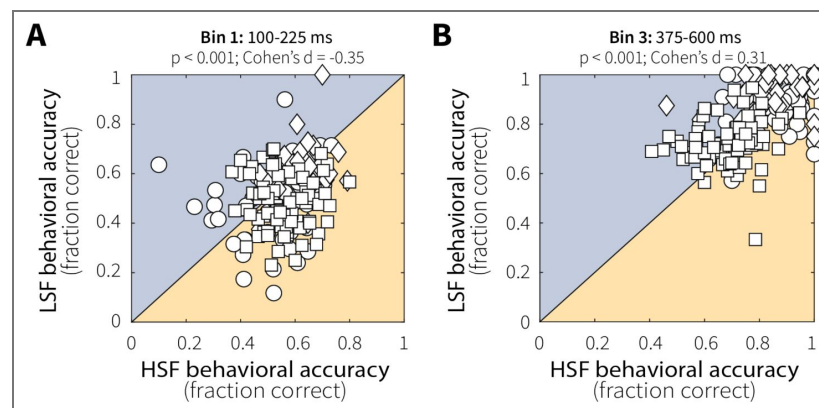
Extended Data Fig. 2. Context-dependent differences in evidence accumulation were consistent across animals.

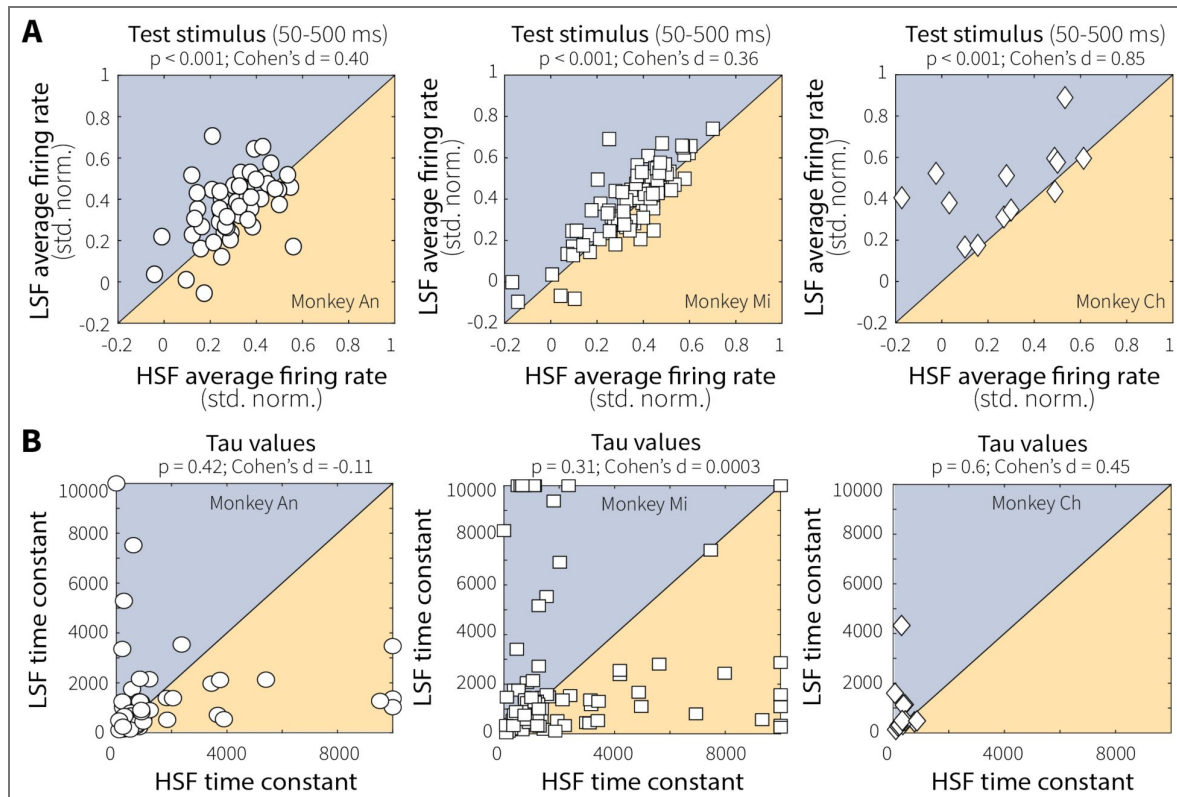
Pairwise comparisons of fitted psychometric slopes for low (LSF) and high (HSF) switch-frequency conditions across sessions for each monkey, as indicated. Slope values were bounded at 0.05, which is roughly the maximum resolvable steepness given our sampling of viewing times. All three monkeys tended to have shallower time-dependent psychometric slopes at HSF relative to LSF (p -values are from a Wilcoxon signed-rank test for equal medians). This result implies that the accuracy of the monkeys' decisions improved more slowly as a function of viewing time at HSF relative to LSF.



Extended Data Fig. 3. Differences in behavioral performance were consistent with context-dependent adjustments in evidence accumulation.

Pairwise comparisons of behavioral accuracy (fraction correct) on switch trials for low (LSF) and high (HSF) switch-frequency conditions, shown separately for shorter-duration (A; Bin 1: 100–225 ms) and longer-duration (B; Bin 3: 375–600 ms) test stimuli. On shorter-duration trials, accuracy was higher in the HSF relative to the LSF condition, whereas at longer viewing durations LSF accuracy surpassed HSF (p -values are from a Wilcoxon signed-rank test for equal medians). This crossover pattern is consistent with adaptive adjustments in evidence accumulation, in which a higher leak rate or lower saturating non-linearity facilitates performance on short switch trials by downweighting irrelevant pre-change evidence but limits performance on longer-duration switch trials as more recent, relevant evidence is also discounted (Glaze et al, 2015).



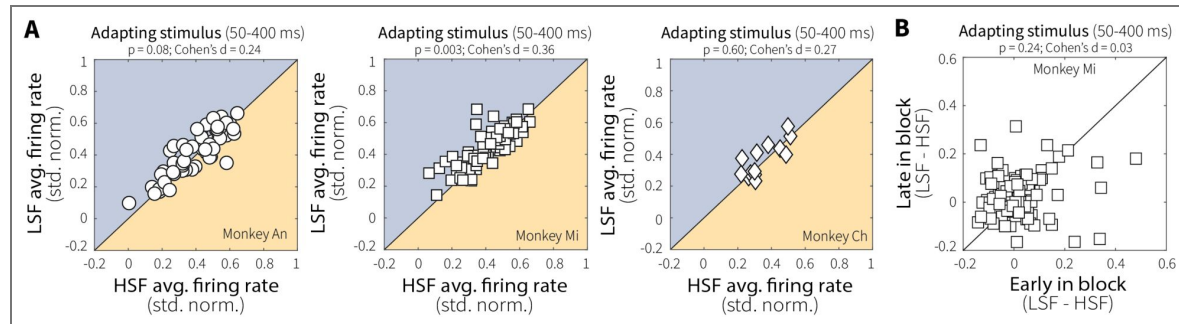


Extended Data Fig. 4. Context-dependent differences in evidence encoding were consistent across animals and reflected changes in response magnitude rather than temporal dynamics.

(A) Pairwise comparisons of mean responses of individual MT units (points) to preferred motion during the test stimulus (50–500 ms after onset) for low (LSF) and high (HSF) switch-frequency conditions across sessions for each monkey, as indicated. All three monkeys exhibited stronger MT responses at LSF relative to HSF (p -values are from a Wilcoxon signed-rank test for equal medians), demonstrating consistent, context-stability-dependent differences in evidence encoding. (B) We quantified the time course of MT neural response by fitting baseline-subtracted, normalized activity during the test stimulus (preferred-motion switch trials only) with a single-exponential function. To facilitate fitting, data for each neuron were divided into three bins (200–330 ms, 340–470 ms, and 480–600 ms following test-stimulus onset) chosen to accommodate initial response latency and include approximately equal trial counts. Fitting was performed separately for low (LSF) and high (HSF) switch-frequency conditions. The exponential time constant (τ , in ms) from best-fitting, single-exponential fits showed no difference between LSF and HSF conditions for each monkey (p -values are from a Wilcoxon signed-rank test for equal medians). Thus, the temporal dynamics of MT responses were not reliably affected by recent temporal statistics.

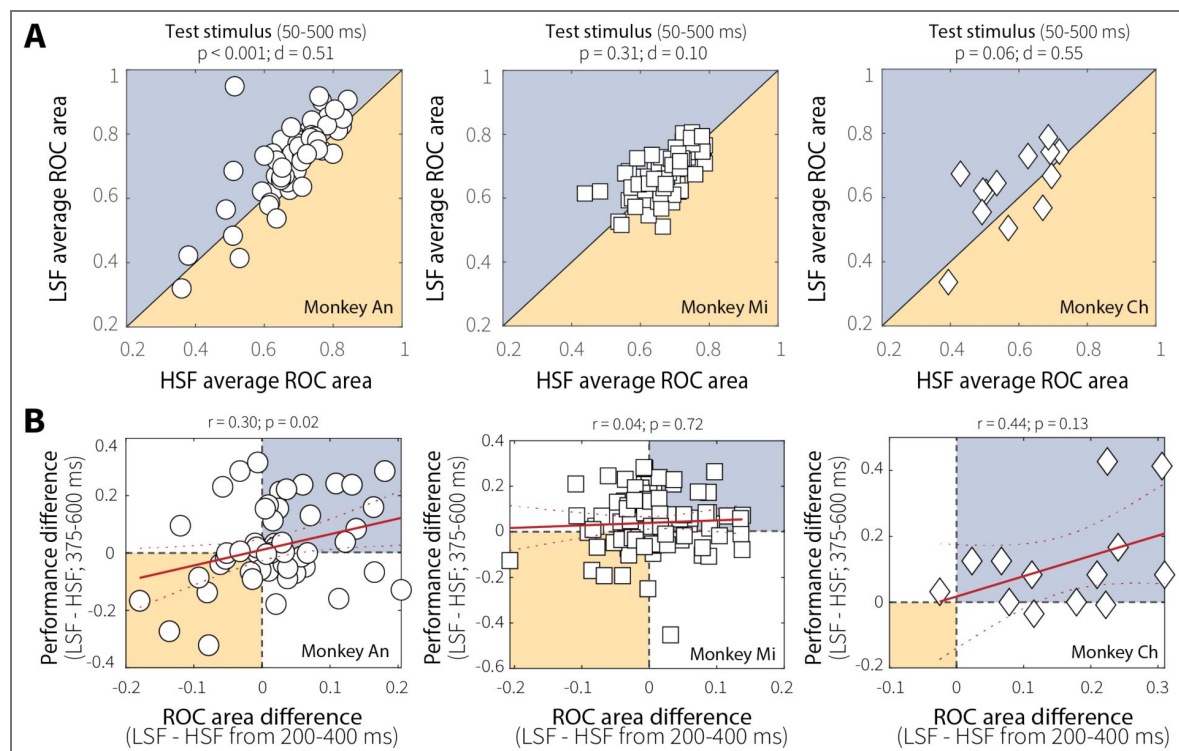
Extended Data Fig. 5. Context-dependent differences in initial MT responses were minimal and stable over the course of the session.

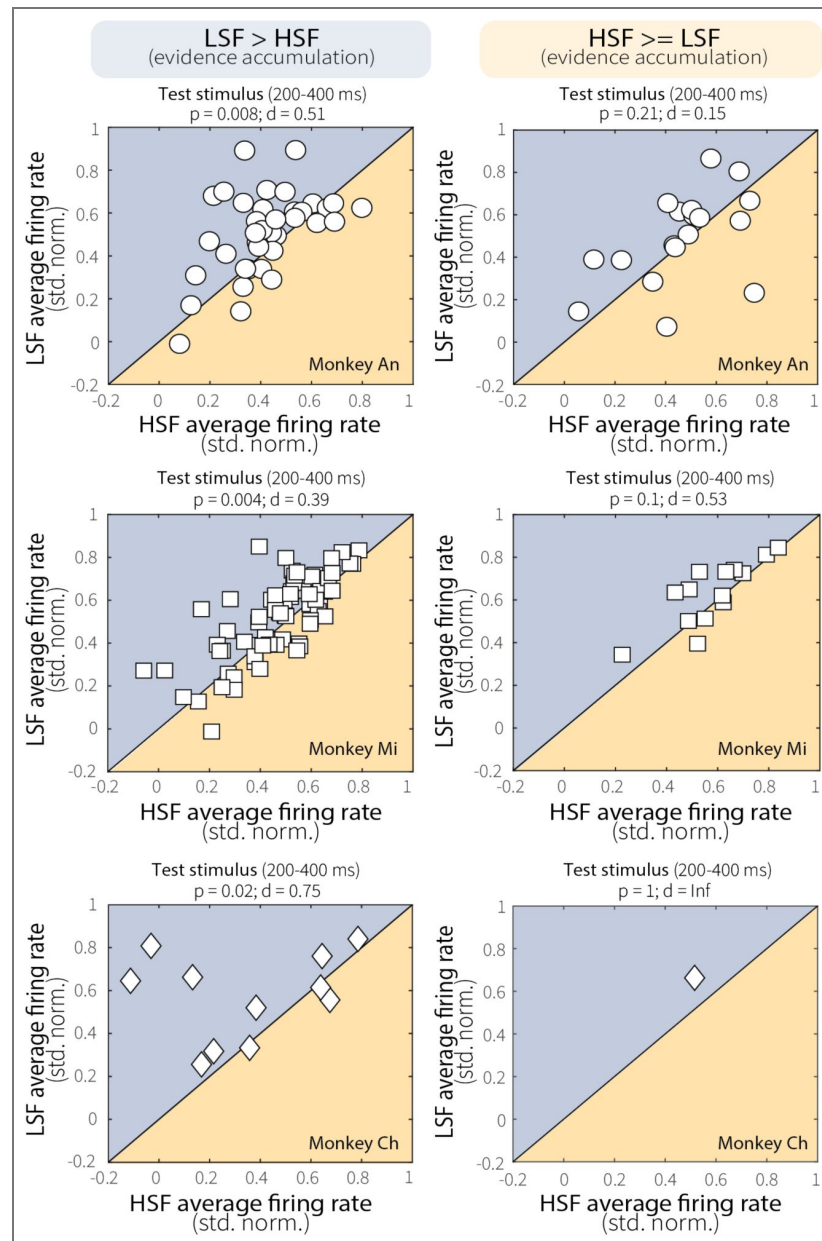
(A) Average MT single-unit responses (points) during onset of the adapting stimulus (50–400 ms) for low (LSF) versus high (HSF) switch-frequency conditions, shown separately for each monkey. Initial responses to preferred motion were matched between LSF and HSF for Monkey An and Monkey Ch but were slightly elevated at LSF for Monkey Mi (p -values are from a Wilcoxon signed-rank test for equal medians). (B) For Monkey Mi, this initial difference in activity between LSF and HSF was stable across early and late blocks of each condition, possibly reflecting expectations about switching dynamics that were learned relatively quickly within each block of trials.



Extended Data Fig. 6. Context-dependent differences in MT evidence discriminability and its relationship to behavioral performance varied across monkeys.

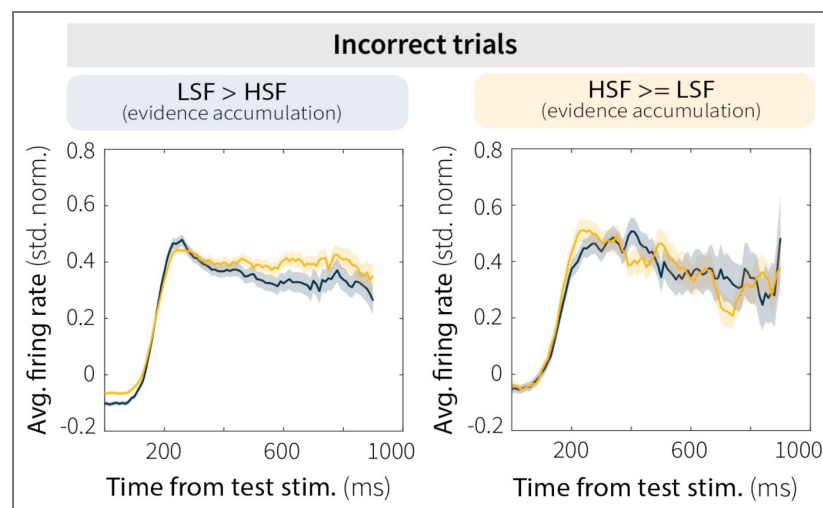
(A) Average ROC area during the stimulus (50–500 ms) for low (LSF) and high (HSF) switch-frequency conditions, shown separately for each monkey. ROC area was greater at LSF relative to HSF for Monkey An (Wilcoxon signed-rank test for equal medians). Monkey Ch showed a similar directional effect with a comparable effect size but for a much smaller sample and thus was not statistically significant. No effect was apparent for Monkey Mi despite a large sample size and context-stability differences in preferred-motion responses during the same stimulus window (Extended Data Fig. 4A). (B) Correlations between context-stability differences in ROC area (LSF-HSF; 200–400 ms) and behavioral performance (LSF-HSF percent correct; 375–600 ms), shown separately for each monkey. A significant positive relationship was observed for Monkey An, with a similar directional trend for Monkey Ch, though power was limited by the smaller number of recorded neurons. No relationship was observed for Monkey Mi despite a large sample size. Red solid and dashed lines indicate linear fits $\pm$ 95% CI.





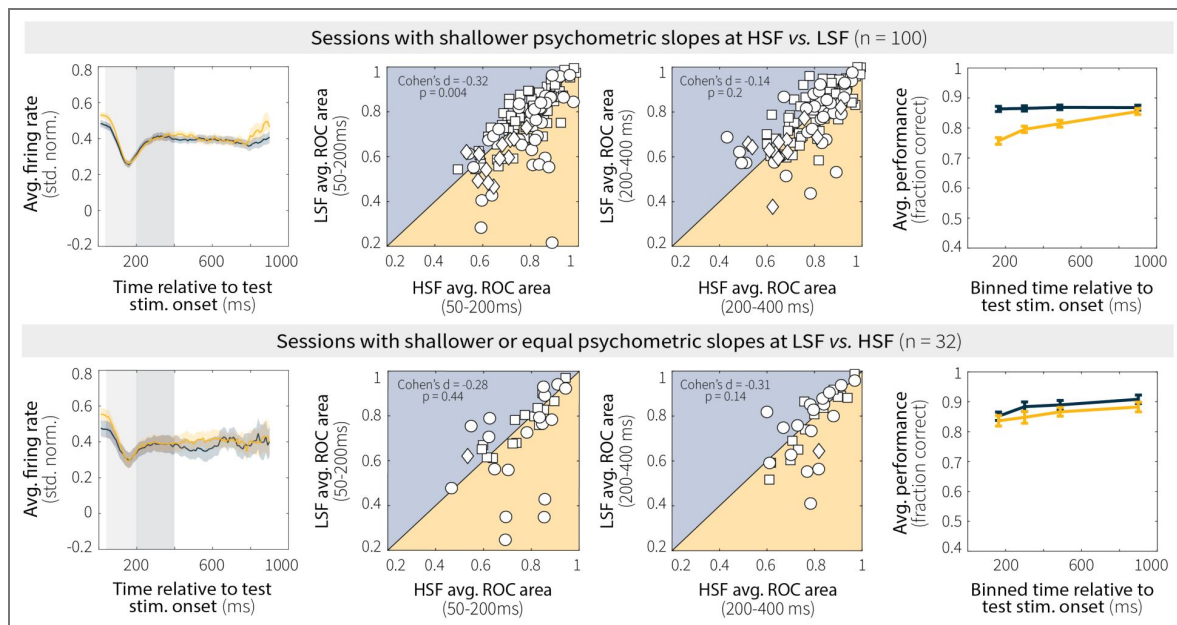
Extended Data Fig. 7. Relationship between MT neural activity and evidence-accumulation behavior was consistent across animals.

Average MT neural activity during the test stimulus (200–400 ms) for low (LSF) and high (HSF) switch-frequency conditions, shown separately for each monkey (rows) and session group (columns). Sessions were divided based on evidence-accumulation behavior: *left column*: sessions in which psychometric slopes were steeper at LSF than HSF ($LSF-HSF > 0$); *right column*: sessions in which psychometric slopes were greater at HSF or equal across conditions ($LSF-HSF \leq 0$). When the monkeys were more sensitive at LSF (left column), MT neurons showed greater activity at LSF relative to HSF. Conversely, when the monkeys were more sensitive at HSF or equally sensitive between conditions (right column), neural responses did not differ between conditions. *p*-values are for a Wilcoxon signed-rank tests for equal medians.



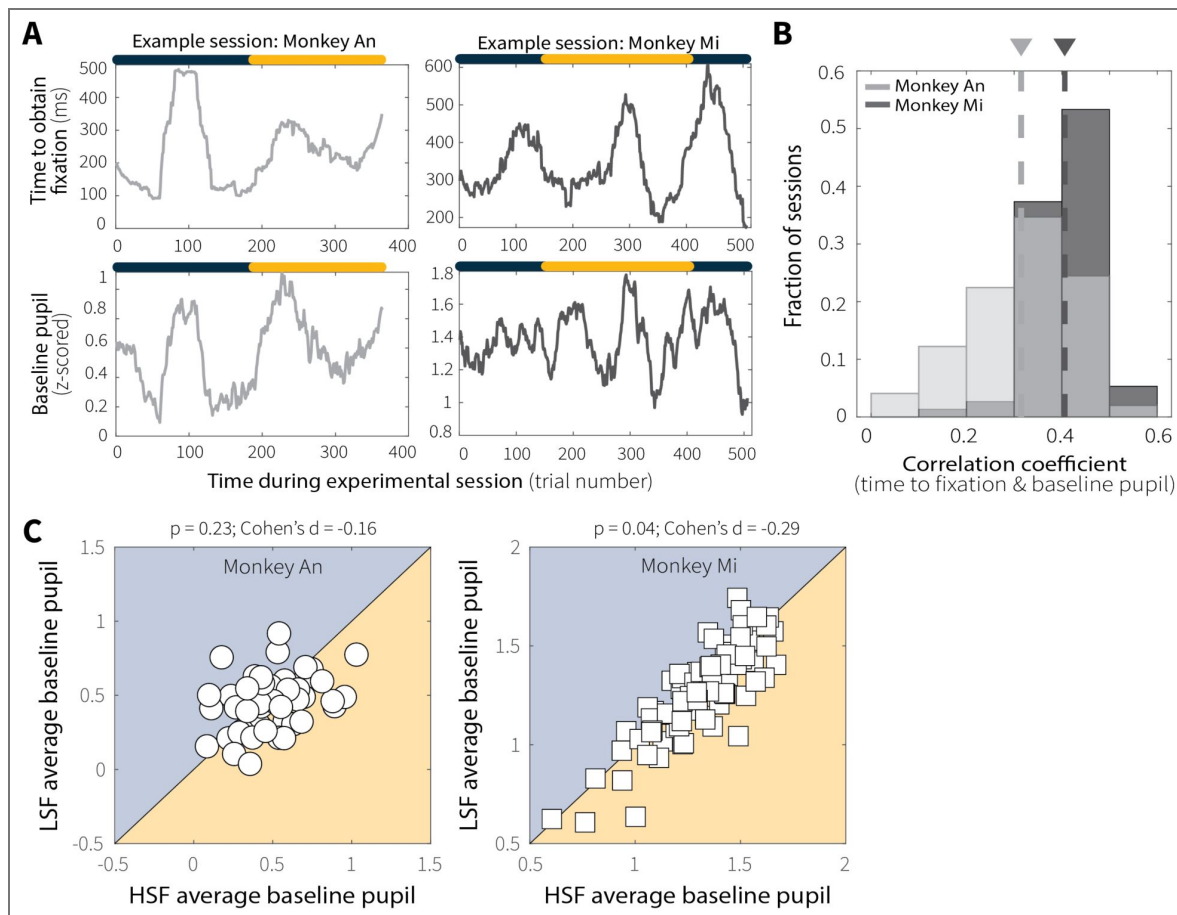
Extended Data Fig. 8. Context-dependent differences in MT activity were specific to correct trials.

Baseline-subtracted and normalized MT firing rates during the test stimulus for low (LSF; blue) and high (HSF; orange) switch-frequency conditions, averaged across incorrect trials for all sessions. Data are shown separately for sessions in which the monkeys were more sensitive (i.e., accumulated more evidence over time) at LSF (left; LSF $n = 1883$, HSF $n = 1939$) or HSF (right; LSF $n = 463$, HSF $n = 395$). Unlike correct trials, error trials showed no consistent modulation by context stability or evidence-accumulation behavior, further confirming that context-dependent changes in MT neural activity were behaviorally relevant.



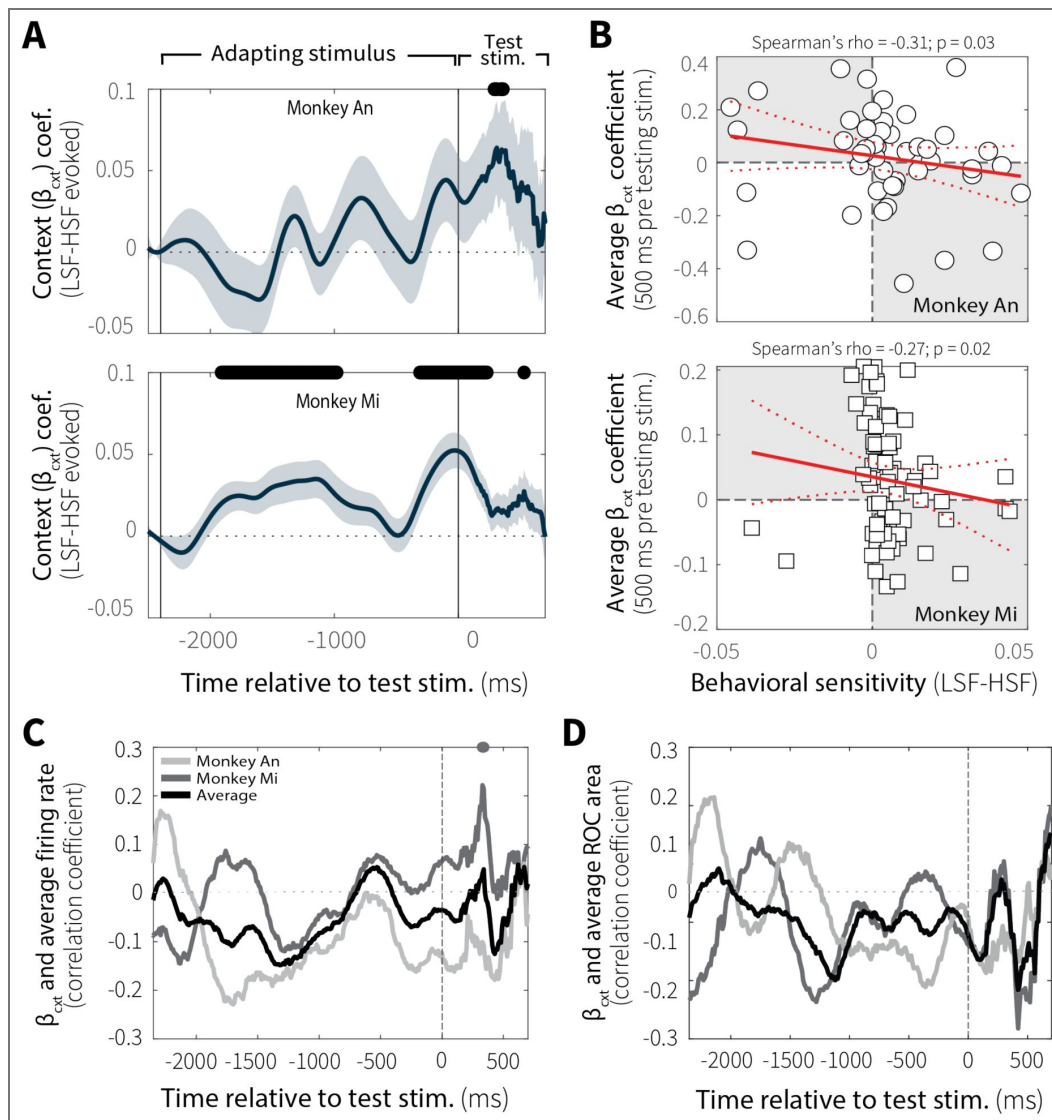
Extended Data Fig. 9. Context-dependent differences in MT neural activity and behavior on non-switch trials.

Data are shown separately for sessions in which psychometric slopes were shallower at HSF vs. LSF (top row; $n = 100$) and psychometric slopes were shallower at LSF vs. HSF or equal across conditions (bottom row; $n = 32$). Columns show: (left) average MT firing rate during the test stimulus, (center-left and center-right) MT evidence discriminability (ROC area) in early (50–200 ms, lighter gray shading in the left panels) and late (200–400 ms, darker gray shading in the left panels) test-stimulus windows, and (right) behavioral performance as a function of test-stimulus duration. In the left panels, the transition from the adapting to the test stimulus, corresponding to a decrease in coherence, is visible as a decrease in MT neural activity ~ 180 ms after test-stimulus onset. Thus, activity in the earlier time bin is likely driven partly by the adapting stimulus, whereas activity in the later time bin is likely driven mainly by the test stimulus. For sessions in which psychometric slopes were shallower at HSF vs. LSF (top row), there was a corresponding context-dependent difference in ROC area in the earlier (center-left), but not later (center-right), time bin. This difference in discriminability was not evident in the other sessions (bottom row, center-left). Consistent with this finding, context-dependent differences in behavioral performance on non-switch trials were also restricted to sessions with shallower psychometric slopes at HSF vs. LSF (top row, right panel). Moreover, in both session groups, behavioral performance on non-switch trials was near (or at) ceiling across viewing durations, reflecting the fact that monkeys accumulated evidence across adapting and test-stimulus epochs on non-switch trials. p -values are from Wilcoxon signed-rank tests for equal medians. Different symbols represent data from different monkeys (Monkey An, circles; Ch, diamonds; Mi, squares).



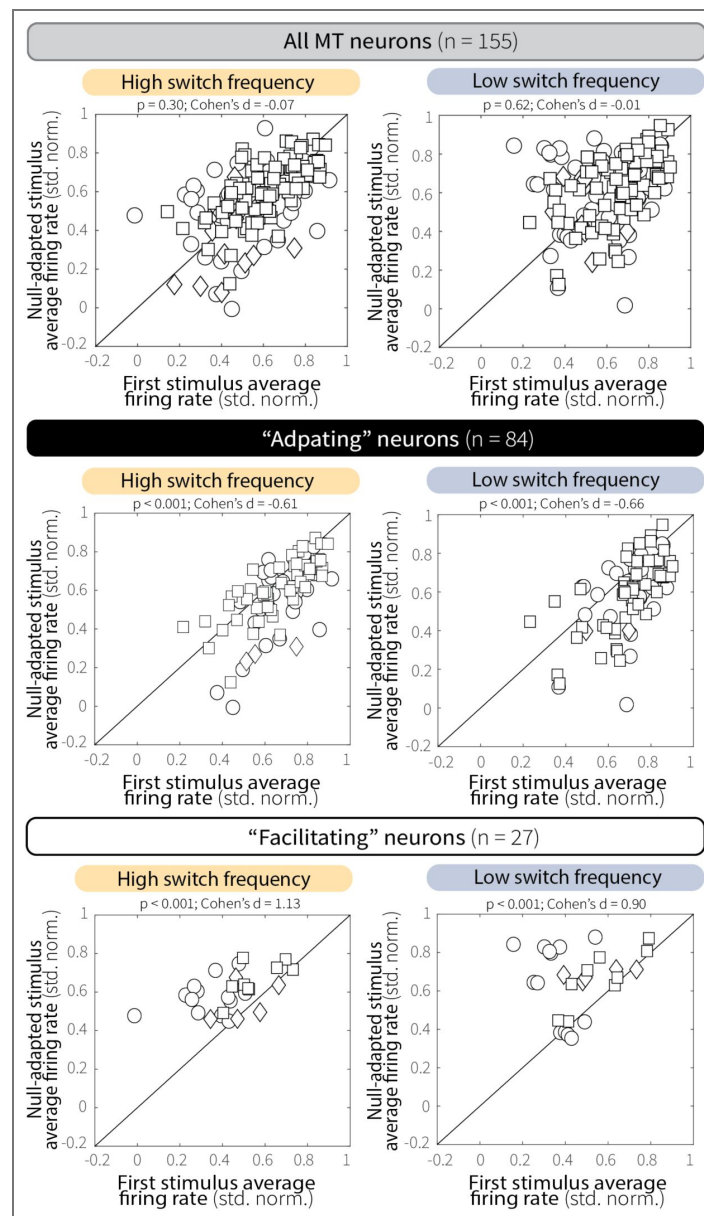
Extended Data Fig. 10. Baseline pupil diameter covaried with fixation-acquisition time across trials, sessions, and monkeys but did not differ across context-stability conditions.

(A) Representative sessions illustrating trial-by-trial relationships between fixation-acquisition time (top) and baseline pupil diameter (bottom) for Monkey An (left, grey) and Monkey Mi (right, black). The colored bar above indicates the active context-stability condition on each trial. (B) Distributions of session-wise Spearman's rank correlation coefficients between fixation-acquisition time and baseline pupil diameter across sessions for Monkey An (grey) and Monkey Mi (black). For both monkeys, the distribution of correlation coefficients was significantly positive (Monkey An: $p < 0.001$, $t = 18.6$, $df = 48$; Monkey Mi: $p < 0.001$, $t = 48.93$, $df = 74$). Mean correlation values are indicated by dashed lines and associated triangles. (C) Session-averaged baseline pupil diameter for LSF versus HSF conditions for Monkey An (left) and Monkey Mi (right). Average baseline pupil size did not reliably differ across context-stability conditions for monkey An and only slightly for monkey Mi (p values are for a Wilcoxon signed-rank test for equal medians).



Extended Data Fig. 11. Evoked pupil diameter related to context-dependent evidence accumulation-behavior, but not MT neural activity.

(A) Context-stability effects on evoked pupil diameter for individual monkeys. Sliding-window regression coefficients (β_{ctx} ; mean $\pm$ SEM across sessions) estimating differences in pupil diameter between low (LSF) and high (HSF) switch-frequency conditions while controlling for baseline pupil size. Black bar (top) indicates windows in which β_{ctx} differed significantly from zero ($p < 0.05$, uncorrected for multiple comparisons). (B) Average evoked pupil differences (β_{ctx} , -500-0 ms relative to test-stimulus onset) plotted versus behavioral sensitivity differences (LSF-HSF psychometric slope) from individual sessions for Monkey An (top) and Monkey Mi (bottom). Red solid and dashed lines indicate a linear fit $\pm$ 95% CI. (C) Spearman's rank correlation between context-stability differences in MT neural activity (LSF-HSF, 50-500 ms) and pupil β_{ctx} coefficients from individual sessions as a function of time relative to test-stimulus onset (computed in 100 ms bins with a 10 ms slide). Individual animals are shown separately (Monkey An: light grey; Monkey Mi: dark grey) along with the across-animal average (black). Corresponding colored bars indicate time points with significant correlations ($p < 0.05$, uncorrected for multiple comparisons). (D) Same analysis for MT discriminability (LSF-HSF ROC area, 50-500 ms) and pupil β_{ctx} coefficients as a function of time relative to test-stimulus.



Extended Data Fig. 12. Effects of null-motion adaptation on preferred-motion responses in MT.

Pairwise comparisons of MT responses to preferred motion before (first stimulus; solid orange line in Fig. 3C) versus after null motion (second, null-adapted stimulus; dashed orange line in Fig. 3C), shown separately for high (HSF, left) and low (LSF, right) switch-frequency conditions. Across the population (top row), prior exposure to null motion did not reliably alter subsequent preferred-motion responses. However, null-motion adaptation produced opposite effects in subgroups of adapting and facilitating neurons. Adapting single units (middle row) showed modest decreases in response to preferred motion following null-motion exposure, whereas facilitating units (bottom row) showed increased responses. Notably, the reduction observed in adapting single units was smaller than that following repeated preferred-motion stimulation. p-values are from Wilcoxon signed-rank tests for equal medians.

Data availability

All data and code have been uploaded to Box and are available at:
<https://upenn.box.com/s/7dx9enb1e1nx19tszrqaur8wrtrflf4>.

Acknowledgements

We thank Jean Zweigle for remarkable and diligent animal care as well as Jafar Bhatti, Long Ding, Catrina Hacker, Victoria Subritzky Katz, and Lowell Thompson for insightful comments on the manuscript. This work was supported by the NIH National Eye Institute (R01 EY015260; JIG), the NIH National Institute of Mental Health (R01 MH127566; JIG), and the National Science Foundation Graduate Research Fellowship Program (NSFGRFP DGE-1845298; KDM).

Additional information

Author contributions

Conceptualization, KDM and JIG; methodology, KDM and JIG; investigation, KDM; visualization, KDM and JIG; funding acquisition, KDM and JIG; project administration, KDM; supervision, JIG; writing – original draft, KDM; writing – review and editing, KDM and JIG.

Funding

Funder	Grant reference number	Author
HHS NIH National Eye Institute (NEI)	R01 EY015260	Joshua I Gold
HHS NIH National Institute of Mental Health (NIMH)	R01 MH127566	Joshua I Gold
NSF National Science Foundation Graduate Research Fellowship Program (GRFP)	NSFGRFP DGE-1845298	Kara D McGaughey

Author ORCID iDs

Joshua I Gold:  <https://orcid.org/0000-0002-6018-0483>

References

- Adibi M, McDonald JS, Clifford CWG, Arabzadeh E (2013) Adaptation Improves Neural Coding Efficiency Despite Increasing Correlations in Variability. *The Journal of Neuroscience* **33**:2108–2120 <https://doi.org/10.1523/JNEUROSCI.3449-12.2013> | PubMed
- Akdoğan B, Balci F, Van Rijn H (2016) Temporal Expectation Indexed by Pupillary Response. *Timing & Time Perception* **4**:354–370 <https://doi.org/10.1163/22134468-00002075>
- Aston-Jones G, Cohen JD (2005) An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience* **28**:403–450 <https://doi.org/10.1146/annurev.neuro.28.061604.135709> | PubMed
- Banerjee A, Parente G, Teutsch J, Lewis C, Voigt FF, Helmchen F (2020) Value-guided remapping of sensory cortex by lateral orbitofrontal cortex. *Nature* **585**:245–250 <https://doi.org/10.1038/s41586-020-2704-z> | PubMed
- Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nature Neuroscience* **10**:1214–1221 <https://doi.org/10.1038/nn1954> | PubMed
- Born RT, Bradley DC (2005) Structure and Function of Visual Area MT. *Annual Review of Neuroscience* **28**:157–189 <https://doi.org/10.1146/annurev.neuro.26.041002.131052> | PubMed
- Brenner N, Bialek W, Steveninck R. de R. van (2000) Adaptive Rescaling Maximizes Information Transmission. *Neuron* **26**:695–702 [https://doi.org/10.1016/S0896-6273\(00\)81205-2](https://doi.org/10.1016/S0896-6273(00)81205-2) | PubMed

- Breton-Provencher V, Sur M** (2019) Active control of arousal by a locus coeruleus GABAergic circuit. *Nature Neuroscience* **22**:218-228 <https://doi.org/10.1038/s41593-018-0305-z> | [PubMed](#)
- Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA** (1996) A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience* **13**:87-100 <https://doi.org/10.1017/S095252380000715X> | [PubMed](#)
- Britten K, Shadlen M, Newsome W, Movshon J** (1992) The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience* **12**:4745-4765 <https://doi.org/10.1523/JNEUROSCI.12-12-04745.1992> | [PubMed](#)
- Cheadle S, Wyart V, Tsetsos K, Myers N, de Gardelle V, Herce Castañón S, Summerfield C** (2014) Adaptive gain control during human perceptual choice. *Neuron* **81**:1429-1441 <https://doi.org/10.1016/j.neuron.2014.01.020> | [PubMed](#)
- Churchland MM, Priebe NJ, Lisberger SG** (2005) Comparison of the Spatial Limits on Direction Selectivity in Visual Areas MT and V1. *Journal of Neurophysiology* **93**:1235-1245 <https://doi.org/10.1152/jn.00767.2004> | [PubMed](#)
- Clarke SE, Longtin A, Maler L** (2015) Contrast coding in the electrosensory system: Parallels with visual computation. *Nature Reviews Neuroscience* **16**:733-744 <https://doi.org/10.1038/nrn4037> | [PubMed](#)
- Cohen-Kashi Malina K, Jubran M, Katz Y, Lampl I** (2013) Imbalance between Excitation and Inhibition in the Somatosensory Cortex Produces Postadaptation Facilitation. *The Journal of Neuroscience* **33**:8463-8471 <https://doi.org/10.1523/JNEUROSCI.4845-12.2013> | [PubMed](#)
- Dalton P** (2000) Psychophysical and Behavioral Characteristics of Olfactory Adaptation. *Chemical Senses* **25**:487-492 <https://doi.org/10.1093/chemse/25.4.487> | [PubMed](#)
- de Gee JW, Colizoli O, Kloosterman NA, Knapen T, Nieuwenhuis S, Donner TH** (2017) Dynamic modulation of decision biases by brainstem arousal systems. *eLife* **6**:e23232 <https://doi.org/10.7554/eLife.23232> | [PubMed](#)
- de Gee JW, Knapen T, Donner TH** (2014) Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences* **111**:E618-E625 <https://doi.org/10.1073/pnas.1317557111> | [PubMed](#)
- Dean I, Harper NS, McAlpine D** (2005) Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience* **8**:1684-1689 <https://doi.org/10.1038/nn1541> | [PubMed](#)
- Derdikman D, Yu C, Haidarliu S, Bagdasarian K, Arieli A, Ahissar E** (2006) Layer-Specific Touch-Dependent Facilitation and Depression in the Somatosensory Cortex during Active Whisking. *Journal of Neuroscience* **26**:9538-9547 <https://doi.org/10.1523/JNEUROSCI.0918-06.2006> | [PubMed](#)
- Devlbiss DM, Page ME, Waterhouse BD** (2006) Locus Coeruleus Regulates Sensory Encoding by Neurons and Networks in Waking Animals. *Journal of Neuroscience* **26**:9860-9872 <https://doi.org/10.1523/JNEUROSCI.1776-06.2006> | [PubMed](#)
- DeWeese M, Zador AM** (1998) Asymmetric Dynamics in Optimal Variance Adaptation. *Neural Computation* **10**:1179-1202 <https://doi.org/10.1162/089976698300017403>
- DeYoe EA, Van Essen DC** (1985) Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature* **317**:58-61 <https://doi.org/10.1038/317058a0> | [PubMed](#)
- Ding L, Gold JI** (2012) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex* **22**:1052-1067 <https://doi.org/10.1093/cercor/bhr178> | [PubMed](#)
- Ding L, Gold JI** (2013) The Basal Ganglia's Contributions to Perceptual Decision Making. *Neuron* **79**:640-649 <https://doi.org/10.1016/j.neuron.2013.07.042> | [PubMed](#)
- Dobler Z, Suresh A, Chari T, Mula S, Tran A, Buonomano DV, Portera-Cailliau C** (2024) Adapting and facilitating responses in mouse somatosensory cortex are dynamic and shaped by experience. *Current Biology* **34**:3506-3521.e5. <https://doi.org/10.1016/j.cub.2024.06.070> | [PubMed](#)

- Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. *Nature* **412**:787-792 <https://doi.org/10.1038/35090500> | PubMed
- Filipowicz AL, Glaze CM, Kable JW, Gold JI (2020) Pupil diameter encodes the idiosyncratic, cognitive complexity of belief updating. *eLife* **9**:e57872 <https://doi.org/10.7554/eLife.57872> | PubMed
- Gilbert CD, Li W (2013) Top-down influences on visual processing. *Nature Reviews Neuroscience* **14**:350-363 <https://doi.org/10.1038/nrn3476> | PubMed
- Glaze CM, Kable JW, Gold JI (2015) Normative evidence accumulation in unpredictable environments. *eLife* **4**:e08825 <https://doi.org/10.7554/eLife.08825> | PubMed
- Gold JI, Stocker AA (2017) Visual Decision-Making in an Uncertain and Dynamic World. *Annual Review of Vision Science* **3**:227-250 <https://doi.org/10.1146/annurev-vision-111815-114511> | PubMed
- Hanks TD, Ditterich J, Shadlen MN (2006) Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience* **9**:682-689 <https://doi.org/10.1038/nn1683> | PubMed
- Higley MJ, Contreras D (2007) Frequency Adaptation Modulates Spatial Integration of Sensory Responses in the Rat Whisker System. *Journal of Neurophysiology* **97**:3819-3824 <https://doi.org/10.1152/jn.00098.2007> | PubMed
- Horowitz GD, Newsome WT (1999) Separate Signals for Target Selection and Movement Specification in the Superior Colliculus. *Science* **284**:1158-1161 <https://doi.org/10.1126/science.284.5417.1158> | PubMed
- Joshi S, Gold JI (2020) Pupil Size as a Window on Neural Substrates of Cognition. *Trends in Cognitive Sciences* **24**:466-480 <https://doi.org/10.1016/j.tics.2020.03.005> | PubMed
- Joshi S, Gold JI (2022) Context-dependent relationships between locus coeruleus firing patterns and coordinated neural activity in the anterior cingulate cortex. *eLife* **11**:e63490 <https://doi.org/10.7554/eLife.63490> | PubMed
- Kastner DB, Baccus SA (2011) Coordinated dynamic encoding in the retina using opposing forms of plasticity. *Nature Neuroscience* **14**:1317-1322 <https://doi.org/10.1038/nn.2906> | PubMed
- Kastner DB, Baccus SA (2013) Spatial segregation of adaptation and predictive sensitization in retinal ganglion cells. *Neuron* **79**:541-554 <https://doi.org/10.1016/j.neuron.2013.06.011> | PubMed
- Keung W, Hagen TA, Wilson RC (2019) Regulation of evidence accumulation by pupil-linked arousal processes. *Nature Human Behaviour* **3**:636-645 <https://doi.org/10.1038/s41562-019-0551-4> | PubMed
- Kheradpezhough E, Adibi M, Arabzadeh E (2017) Response dynamics of rat barrel cortex neurons to repeated sensory stimulation. *Scientific Reports* **7**:11445 <https://doi.org/10.1038/s41598-017-11477-6> | PubMed
- Kim JN, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience* **2**:176-185 <https://doi.org/10.1038/5739> | PubMed
- Kohn A, Movshon JA (2003) Neuronal Adaptation to Visual Motion in Area MT of the Macaque. *Neuron* **39**:681-691 [https://doi.org/10.1016/S0896-6273\(03\)00438-0](https://doi.org/10.1016/S0896-6273(03)00438-0) | PubMed
- Krishnamurthy K, Nassar MR, Sarode S, Gold JI (2017) Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nature Human Behaviour* **1**:0107 <https://doi.org/10.1038/s41562-017-0107> | PubMed
- Lamp I, Katz Y (2017) Neuronal adaptation in the somatosensory system of rodents. *Neuroscience* **343**:66-76 <https://doi.org/10.1016/j.neuroscience.2016.11.043> | PubMed
- Laughlin SB (1989) The role of sensory adaptation in the retina. *Journal of Experimental Biology* **146**:39-62 <https://doi.org/10.1242/jeb.146.1.39> | PubMed
- Law C.-T, Gold JI (2008) Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience* **11**:505-513 <https://doi.org/10.1038/nn2070> | PubMed

- Lesica NA, Jin J, Weng C, Yeh C-I, Butts DA, Stanley GB, Alonso J-M (2007) Adaptation to Stimulus Contrast and Correlations during Natural Visual Stimulation. *Neuron* **55**:479-491 <https://doi.org/10.1016/j.neuron.2007.07.013> | PubMed
- Lovejoy LP, Krauzlis RJ (2010) Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. *Nature Neuroscience* **13**:261-266 <https://doi.org/10.1038/nn.2470> | PubMed
- Mathôt S, Fabius J, Van Heusden E, Van der Stigchel S (2018) Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods* **50**:94-106 <https://doi.org/10.3758/s13428-017-1007-2> | PubMed
- McGuire JT, Nassar MR, Gold JI, Kable JW (2014) Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron* **84**:870-881 <https://doi.org/10.1016/j.neuron.2014.10.013> | PubMed
- Młynarski WF, Hermundstad AM (2021) Efficient and adaptive sensory codes. *Nature Neuroscience* **24**:998-1009 <https://doi.org/10.1038/s41593-021-00846-0> | PubMed
- Movshon JA, Newsome WT (1996) Visual Response Properties of Striate Cortical Neurons Projecting to Area MT in Macaque Monkeys. *Journal of Neuroscience* **16**:7733-7741 <https://doi.org/10.1523/JNEUROSCI.16-23-07733.1996> | PubMed
- Murphy PR, Vandekerckhove J, Nieuwenhuis S (2014) Pupil-Linked Arousal Determines Variability in Perceptual Decision Making. *PLOS Computational Biology* **10**:e1003854 <https://doi.org/10.1371/journal.pcbi.1003854> | PubMed
- Murphy PR, Wilming N, Hernandez-Bocanegra DC, Prat-Ortega G, Donner TH (2021) Adaptive circuit dynamics across human cortex during evidence accumulation in changing environments. *Nature Neuroscience* **24**:987-997 <https://doi.org/10.1038/s41593-021-00839-z> | PubMed
- Nagel KI, Doupe AJ (2006) Temporal Processing and Adaptation in the Songbird Auditory Forebrain. *Neuron* **51**:845-859 <https://doi.org/10.1016/j.neuron.2006.08.030> | PubMed
- Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience* **15**:1040-1046 <https://doi.org/10.1038/nn.3130> | PubMed
- Navarra RL, Clark BD, Gargiulo AT, Waterhouse BD (2017) Methylphenidate Enhances Early-Stage Sensory Processing and Rodent Performance of a Visual Signal Detection Task. *Neuropsychopharmacology* **42**:1326-1337 <https://doi.org/10.1038/npp.2016.267> | PubMed
- Newsome W, Pare E (1988) A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *The Journal of Neuroscience* **8**:2201-2211 <https://doi.org/10.1523/JNEUROSCI.08-06-02201.1988> | PubMed
- O'Reilly JX, Schüffelen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS (2013) Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences* **110**:E3660-E3669 <https://doi.org/10.1073/pnas.1305373110> | PubMed
- Piet AT, El Hady A, Brody CD (2018) Rats adopt the optimal timescale for evidence integration in a dynamic environment. *Nature Communications* **9**:4265 <https://doi.org/10.1038/s41467-018-06561-y> | PubMed
- Rodman HR, Gross CG, Albright TD (1989) Afferent basis of visual response properties in area MT of the macaque. I. Effects of striate cortex removal. *Journal of Neuroscience* **9**:2033-2050 <https://doi.org/10.1523/JNEUROSCI.09-06-02033.1989> | PubMed
- Salzman CD, Britten KH, Newsome WT (1990) Cortical microstimulation influences perceptual judgements of motion direction. *Nature* **346**:174-177 <https://doi.org/10.1038/346174a0> | PubMed
- Schwarz LA, Luo L (2015) Organization of the locus coeruleus-norepinephrine system. *Current Biology: CB* **25**:R1051-R1056 <https://doi.org/10.1016/j.cub.2015.09.039> | PubMed

- Schwiedrzik CM, Sudmann SS (2020) Pupil Diameter Tracks Statistical Structure in the Environment to Increase Visual Sensitivity. *Journal of Neuroscience* **40**:4565-4575 <https://doi.org/10.1523/JNEUROSCI.0216-20.2020> | PubMed
- Seay MJ, Natan RG, Geffen MN, Buonomano DV (2020) Differential Short-Term Plasticity of PV and SST Neurons Accounts for Adaptation and Facilitation of Cortical Neurons to Auditory Tones. *Journal of Neuroscience* **40**:9224-9235 <https://doi.org/10.1523/JNEUROSCI.0686-20.2020> | PubMed
- Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology* **86**:1916-1936 <https://doi.org/10.1152/jn.2001.86.4.1916> | PubMed
- Solomon SG, Kohn A (2014) Moving Sensory Adaptation beyond Suppressive Effects in Single Neurons. *Current Biology* **24**:R1012-R1022 <https://doi.org/10.1016/j.cub.2014.09.001> | PubMed
- Tikidji-Hamburyan A, Reinhard K, Seitter H, Hovhannisyan A, Procyk CA, Allen AE, Schenk M, Lucas RJ, Münch TA (2015) Retinal output changes qualitatively with every change in ambient illuminance. *Nature Neuroscience* **18**:66-74 <https://doi.org/10.1038/nn.3891> | PubMed
- Tolias AS, Smirnakis SM, Augath MA, Trinath T, Logothetis NK (2001) Motion Processing in the Macaque: Revisited with Functional Magnetic Resonance Imaging. *Journal of Neuroscience* **21**:8594-8601 <https://doi.org/10.1523/JNEUROSCI.21-21-08594.2001> | PubMed
- Ulanovsky N, Las L, Farkas D, Nelken I (2004) Multiple Time Scales of Adaptation in Auditory Cortex Neurons. *Journal of Neuroscience* **24**:10440-10453 <https://doi.org/10.1523/JNEUROSCI.1905-04.2004> | PubMed
- Usher M, McClelland JL (2001) The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review* **108**:550-592 <https://doi.org/10.1037/0033-295X.108.3.550> | PubMed
- Van Wezel RJA, Britten KH (2002) Motion Adaptation in Area MT. *Journal of Neurophysiology* **88**:3469-3476 <https://doi.org/10.1152/jn.00276.2002>
- Veliz-Cuba A, Kilpatrick Z, Josić K (2016) Stochastic Models of Evidence Accumulation in Changing Environments. *SIAM Review* **58** <https://doi.org/10.1137/15m1028443>
- Vinck M, Batista-Brito R, Knoblich U, Cardin JA (2015) Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron* **86**:740-754 <https://doi.org/10.1016/j.neuron.2015.03.028> | PubMed
- Waiblinger C, McDonnell ME, Reedy AR, Borden PY, Stanley GB (2022) Emerging experience-dependent dynamics in primary somatosensory cortex reflect behavioral adaptation. *Nature Communications* **13**:534 <https://doi.org/10.1038/s41467-022-28193-z> | PubMed
- Waiblinger C, Reedy AR, Stanley GB (2025) An adaptive and flexible role for primary sensory cortex. *Nature Neuroscience* 1-11 <https://doi.org/10.1038/s41593-025-02124-9> | PubMed
- Wark B, Fairhall A, Rieke F (2009) Timescales of Inference in Visual Adaptation. *Neuron* **61**:750-761 <https://doi.org/10.1016/j.neuron.2009.01.019> | PubMed
- Wark B, Lundstrom BN, Fairhall A (2007) Sensory adaptation. *Current Opinion in Neurobiology* **17**:423-429 <https://doi.org/10.1016/j.conb.2007.07.001> | PubMed
- Waterhouse BD, Moises HC, Woodward DJ (1998) Phasic activation of the locus coeruleus enhances responses of primary sensory cortical neurons to peripheral receptive field stimulation. *Brain Research* **790**:33-44 [https://doi.org/10.1016/S0006-8993\(98\)00117-6](https://doi.org/10.1016/S0006-8993(98)00117-6) | PubMed
- Weber AI, Krishnamurthy K, Fairhall AL (2019) Coding Principles in Adaptation. *Annual Review of Vision Science* **5**:427-449 <https://doi.org/10.1146/annurev-vision-091718-014818> | PubMed
- Whitmire CJ, Stanley GB (2016) Rapid Sensory Adaptation Redux: A Circuit Perspective. *Neuron* **92**:298-315 <https://doi.org/10.1016/j.neuron.2016.09.046> | PubMed
- Wyrick D, Mazzucato L (2021) State-Dependent Regulation of Cortical Processing Speed via Gain Modulation. *Journal of Neuroscience* **41**:3988-4005 <https://doi.org/10.1523/JNEUROSCI.1895-20.2021> | PubMed

Peer reviews

Reviewer #1 (Public review):

McGaughey and Gold ask where in the decision process the flexibility of evidence accumulation arises, proposing that it is not solely a property of downstream integrators but is also supported by stimulus-specific sensory adaptation in the middle temporal area (MT). Recording single-unit activity in rhesus macaques during a motion direction-discrimination task in which an adapting stimulus of varying temporal stability precedes an identical test stimulus, they find that more rapidly changing contexts produce weaker and less discriminable MT responses to the test stimulus, which they argue accounts in part for context-dependent changes in decision-making behavior. Through session-level correlations they further identify pupil-linked arousal as a parallel, apparently separable contributor.

The main strength is the shift of perspective toward the encoding stage: rather than treating MT as a static input to flexible downstream integrators, the authors show that early sensory cortex can itself contribute adaptive, context-dependent signals that shape behavior. The conceptual advance is supported by a well-designed paradigm-total exposure to each motion direction is matched across conditions and the test stimulus is held identical-together with single-unit recordings and simultaneous pupillometry. The behavioral effect is consistent across three animals, and the fact that context-dependent differences emerge over repeated stimulus presentations within a trial, rather than as a sustained baseline offset across blocks, ties the effect convincingly to stimulus-specific adaptation.

The behavioral effect constrains the temporal dynamics of decision formation but does not uniquely identify its algorithmic basis: a leak, a saturating non-linearity, or a reduction in the gain of integration are all compatible with a shallower rise of accuracy with viewing time, and the reduced MT discriminability is itself an encoding-stage efficiency effect of this kind. The manuscript appropriately treats the algorithmic basis as unresolved, noting that distinguishing these accounts would require analyses not available here, such as reverse-correlation or motion-energy kernels with lower-coherence test stimuli.

The inference that the adaptation- and arousal-related signals operate independently rests on the absence of session-wise correlations between the neural and pupil measures and their behavioral contributions. Given the noise in the trial-wise estimates, this is best read as consistent with, rather than demonstrating, true independence, as the authors note.

Overall, the authors largely achieve their aim of showing that sensory adaptation in MT shapes the evidence available for time-dependent perceptual decisions. The evidence for a sensory-encoding contribution is convincing, while the claim of independence between adaptation and arousal is more tentative and is framed as such.

<https://doi.org/10.7554/eLife.110685.2.sa2>

Reviewer #2 (Public review):

McGaughey and Gold trained rhesus macaque monkeys to perform a motion-direction discrimination task in which a behaviorally irrelevant adapting stimulus with either fast or slow direction alternations preceded a variable-duration test stimulus, while simultaneously recording single-unit activity in area MT and pupil diameter. They report that adaptation to the more rapidly changing stimulus was associated with reduced behavioral sensitivity, attenuated test-evoked MT responses, and larger pupil-linked arousal signals. The authors interpret these behavioral changes as evidence for context-dependent adjustments to the temporal dynamics of decision formation and argue that these adjustments are supported by both sensory adaptation in MT and arousal-related mechanisms. More broadly, they conclude

that flexible evidence accumulation in dynamic environments arises from distributed adjustments across sensory encoding and neuromodulatory systems rather than solely from changes within a downstream accumulator. If correct, this interpretation has important implications not only for our understanding of perceptual decision making, but also for broader theories concerning the functional role of sensory adaptation.

The conclusions of the paper are generally supported by the data. Evidence for adaptation-induced changes in sensory encoding, behavior, and pupil dynamics is convincing, and the revised manuscript substantially strengthens the connection between the behavioral findings and the proposed decision-making framework.

Comments on revised version.

The revised manuscript provides a clearer account of how recent stimulus history influences behavioral performance. In the original version, aspects of the psychometric functions were interpreted as evidence for a more leaky evidence-accumulation process, although some of these effects could potentially have reflected alternative mechanisms, including influences of the adapting stimulus on short-duration trials. The additional analyses and discussion included in the revision clarify that information from the adapting stimulus contributes to behavior at short viewing durations and appropriately temper claims regarding the specific computational mechanism underlying the observed behavioral effects. While the data do not uniquely identify whether these effects arise from changes in leak, other nonlinearities, or related decision processes, they provide convincing evidence that recent temporal context influences the temporal dynamics of decision formation.

My original review also noted that different sections of the manuscript relied on different behavioral metrics and analytical approaches when relating behavioral changes to neural and pupil-linked measures. The revised manuscript now provides a clearer rationale for these choices, including distinctions arising from the different trial types and time windows used in the neural and pupil analyses.

<https://doi.org/10.7554/eLife.110685.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Reviewer #1 (Public review):

(1) Alternative mechanisms for performance differences.

The authors assume that the difference in performance between the low-switch (LS) and high-switch (HS) frequency conditions is explained by a change in the "leakiness" of integration. However, several other mechanisms could potentially explain this effect:

(1) Temporal Uncertainty: Integration might start later in the HS condition, leading to lower performance.

(2) Reduced Efficiency: Integration could be less efficient in the HS condition (i.e., lower signal-to-noise ratio) without a change in the leak parameter itself.

(3) Evidence Contamination: Motion information from the adapting stimulus in the HS condition may be integrated rather than ignored, which might be the case since the transition from the adapting to the test stimulus is not externally cued.

To distinguish between these alternatives, I suggest two possible analyses. First, a formal model comparison could be performed, though I acknowledge this may be inconclusive in the absence of response-time data. Second, an analysis of motion energy kernels could

be revealing; the leak hypothesis makes the specific prediction that for long test stimuli, early samples should contribute more to the choice in the LS condition than in the HS condition, relative to late samples.

We thank the reviewer for raising these important points. We agree that we cannot definitively identify the algorithmic underpinnings of the behavioral effects we report and have made substantial revisions to the manuscript to be clearer about what is supported and what is speculative in our claims. Most importantly, we agree that we do not know if the context-dependent differences in how accuracy depends on viewing time are based on adjustments to a leak or to something else (e.g., a saturating non-linearity, as we identified in Glaze et al, 2015, that is separate from the leak itself), which we cannot resolve with this dataset, even with more formal model comparisons. We therefore:

Changed the wording throughout the manuscript to refer to changes in leakiness as just one of several possible sources of the behavioral differences. We also added this point to the list of “limitations” (and possible future directions, including using motion-energy kernels, which would require us to use lower-coherence test stimuli) in the Discussion (L487-493).

Added a new figure panel (Fig. 2D), a new Extended Data figure (Extended Data Fig. 3), and additional explanatory text (L168-175) that collectively describe the behavior in more detail, including quantifying a “crossover” dynamic similar to what we reported previously (Glaze et al, 2015).

Added new explanations (L152-163) and analyses (Extended Data Fig. 9) indicating that the monkeys used some information from the end of the adapting stimulus to inform their decisions, which accounts for the patterns of choices at the shortest viewing durations.

Indicate that the context-dependent differences in the slopes of the psychometric functions (and complementary analyses based on “raw” accuracy measures as a function of binned viewing duration) rule out the temporal uncertainty and evidence contamination explanations, but are consistent with effects on the temporal dynamics of the decision process (L175-179).

(2) Independence of neural and pupil-linked signals.

The authors take the lack of session-wise correlation between context-dependent contributions from neural and pupil terms as evidence that these two signals provide independent contributions to the behavioral effect. However, could this lack of correlation simply be a result of high variability or noise in these estimates? The data shown in Figure 7B suggests that measurements are very noisy, which might obscure a potential relationship.

We agree that the lack of session-wise correlation between neural and pupil terms cannot be taken as definitive evidence of independence. We have both softened the language around the claim (L368) and added a sentence to the Discussion (L464-468) acknowledging that this lack of correlation may reflect underlying noise and/or variability rather than true independence of the underlying mechanisms.

Reviewer #1 (Recommendations for the authors):

(3) The neural data analyses rely fundamentally on “switch” trials (Figures 3-5). It might be informative to also examine “non-switch” trials to see if there are specific neural markers indicating the exact moment the motion stimulus becomes behaviorally relevant. Given that this may fall outside the primary focus of the paper, it is up to the authors whether to pursue this line of inquiry.

We thank the reviewer for this suggestion. We agree and have added new analyses of data from non-switch trials (Extended Data Fig. 9), which show some effects of stimulus information from the adapting epoch on the monkeys' choices, as we detail below in response to related comments from the other reviewers.

Reviewer #2 (Public review):

Aspects of the behavioral analysis would benefit from a tighter connection between theoretical claims about evidence accumulation and the empirical features of the psychometric functions. For example, the rightward shifts observed across adapting conditions are interpreted as consistent with a reset of accumulation on switch trials, but similar patterns could also arise from failures to detect the test stimulus on a subset of trials, leading responses to default to the final adaptor direction. Likewise, changes in psychometric slope and asymptote are attributed to differences in evidence accumulation without explicit modelling or consideration of alternative explanations.

Clarifying how specific features of the psychometric functions map onto distinct components of the decision process will strengthen the link between the theoretical framework and the behavioral data.

We agree and have made substantial revisions to address these important points. Specifically, we added a new figure panel (Fig. 2D), new Extended Data Figures (3 and 9), and several lines of explanatory text (L152-179) that collectively describe the behavior in more detail, including clarifying that: 1) for the shortest viewing durations, the monkeys' decisions were informed by information from the adapting stimulus, which accounts for generally lower accuracy on LSF (longer exposure to the final adapting direction, thus more accumulated evidence for that direction before processing the switch) vs. HSF (shorter exposure to the final adapting direction, thus less accumulated evidence for that direction before processing the switch) switch trials; and 2) as viewing duration increased, the rate of rise of accuracy versus viewing duration was higher for LSF vs. HSF trials, implying differences in the process of evidence accumulation. As detailed in our response to a similar comment from Reviewer 1, above, we are now careful to temper our claims about the specific computational basis (e.g., a leak or other form of nonlinearity) for these differences.

We also de-emphasized our treatment of the asymptotes of the psychometric functions. In principle, these regimes could give insights into leakiness (which can limit the total amount of information that can be accumulated) and lapses (which are measured at the asymptotes). In practice, however, the long-duration trials that constitute the asymptotes were relatively under sampled (to promote the unpredictability of the offset of the stimulus, which we believed was the more important consideration when designing the experiment), yielding unreliable estimates.

A slight concern is the lack of a consistent analytical approach for relating behavioral changes to neural and pupil-linked measures. Different sections of the manuscript rely on different behavioral metrics-such as differences in accuracy within a selected stimulus-duration range (e.g., Figure 5C) or psychometric slope differences (Figure 6C) without clear justification for these choices. The analytical approach likewise varies between simple correlational analyses (Figure 5C, Figure 6C), pseudo-experimental group comparisons (Figures 5D, E), and the inclusion of neural or pupil terms in the behavioral psychometric regression model (Figure 7B). While each metric and approach may be defensible in isolation, adopting a more consistent framework will help convince readers that the reported effects are robust and not contingent on the selective choice of metric or analysis.

We thank the reviewer for this thoughtful critique and agree that the rationale for our choice

of behavioral metrics and analytical approaches could be stated more clearly. We have added text to the relevant sections of the Results (L247-251) clarifying these choices. In particular:

The neural analyses (Figures 3D-E, Figure 4, Figure 5D-E) focused on preferred-motion switch trials, because: 1) low switch-frequency non-switch trials provide an additional 800 ms of exposure to the final adapting-stimulus motion direction relative to high switch-frequency non-switch trials, which confounds comparisons of context-dependent evidence encoding between conditions, and 2) MT neurons exhibit minimal responses to null motion (although note that we also included analyses based on ROC area, which is computed from both preferred- and null-motion switch trials, to account for possible contributions of null-motion responses; Figure 5A-C). Thus, to ensure a meaningful comparison between neural and behavioral measures, we used behavioral accuracy on switch trials as the relevant metric in Figure 5C-E, rather than psychometric slope, which is estimated across both switch and non-switch trials.

The pupil analyses (Figure 6) focused on a time window preceding test-stimulus onset, representing the arousal state around when the decision process started, and included both switch and non-switch trials. Thus, for these analyses we used psychometric slope, which is estimated across both switch and non-switch trials.

We used several different analyses to compare and contrast the neural-behavioral and pupil-behavioral relationships because they provide complementary and useful insights. The correlational analyses in Figures 5C and 6C characterize session-level relationships between neural/pupil signals and behavior. The group comparisons in Figures 5D-E provide a complementary visualization of the same relationship. The model-based approach in Figure 7 then allows direct quantification of the trial-wise contributions of each signal to behavior within a common framework. Importantly, the conclusions drawn from each approach converge on the same interpretation, which we believe speaks to the robustness of the reported effects.

Reviewer #2 (Recommendations for the authors):

(1) Figure 2 legend. Description of 'running average (5-trial window)' is unclear - presumably this is a running average in stimulus space rather than across trials.

We thank the reviewer for flagging this ambiguity. We have updated the legend (L136-137) to clarify that the running average is computed across trials sorted by test-stimulus duration.

(2) L158. Difficult to establish an asymptotic performance level for HSF conditions within the stimulus duration range tested.

We have removed the reference to asymptotic performance and replaced it with a discussion of performance on longer-duration switch trials in the context of the newly added Figure 2D.

(3) L515 Equation 1. While this is a standard formulation of lapse rate in psychometric functions, the construction here in terms of switch probability is not standard. Given the task and training, it seems more likely that on lapse trials, the animal will respond according to the last adapted direction (rather than randomly switch/stay with equal probability).

We thank the reviewer for this point. We agree that it is possible that on at least some of the “lapse” trials the monkeys may respond according to the final adapting-stimulus direction rather than choosing randomly. However, we cannot distinguish those alternatives using this task design. We include a statement to this effect in Methods (L569-571).

To explore the idea further, we refit the behavioral data using separate upper and lower asymptotes corresponding to lapse rates on switch and non-switch trials, respectively. Across

monkeys, there were no significant differences between upper and lower lapse rates for either low (Wilcoxon signed-rank test for equal medians: $p = 0.15$, Cohen's $d = -0.13$) or high switch frequency ($p = 0.07$, Cohen's $d = -0.16$) conditions. So, at the very least, there was no evidence for lapse-like errors driven by switch- (or non-switch-) specific defaults to the final adapting direction.

(4) L256. *Statistical significance of attenuation is not directly tested here.*

We have replaced "were attenuated" with "we did not identify any reliable context-stability differences" (L297) to accurately reflect what was directly tested without implying a statistical comparison between groups of sessions that was not performed.

(5) L429. *Does the increase in explanatory power warrant the increased complexity of the model here?*

We thank the reviewer for raising this important point. We used Tjur's pseudo- R^2 because it does not increase by default with added model complexity, making it more conservative than other R^2 measures in this respect. Tjur's pseudo- R^2 is a coefficient of discrimination, and as such its value increases only when additional terms improve the model's ability to separate predicted probabilities across response outcomes. Thus, the observed increases in explanatory power when adding neural or pupil terms reflect real improvements in discriminability rather than an artifact of model complexity. We have added a brief clarification of this point to the Methods (L662-664).

Reviewer #3 (Public review):

The task design may not be optimal. While the amount of time the monkey is exposed to each motion direction during the adapting stimulus is matched, it's hard to know if the reduced MT responses to the test stimulus are truly due to the greater frequency of switches during the HSF adapting stimulus or because the monkeys have been exposed to more repetitions of the stimulus. It's increased sensory adaptation in either case, but it makes it problematic to interpret this as temporal context-dependent adaptation specifically. I think this could potentially be partially addressed by an analysis that is in the paper, but could potentially be emphasized/fleshed out more, specifically the results shown in Figure 4D that seem to show that most of the reduction in neural response for adapting units occurs between the first and second stimuli.

The reviewer raises an important point. The number of stimulus repetitions and switch frequency are confounded in the experimental design, making it difficult to attribute context-dependent differences in MT responses to the temporal pattern of switches rather than to accumulated repetitions. We also note, as the reviewer acknowledges, the observed differences reflect sensory adaptation either way. Figure 4D does offer relevant evidence, suggesting that a majority of the change in neural response occurred with just one stimulus repetition. This finding complicates an interpretation where adaptation scales with the number of stimulus repetitions. We have added several lines to the Results about these points (L231-233).

The pupillometric analysis seems to be an indirect way of assessing whether the accumulator itself might be modulated by temporal context, but the link could be made clearer. The authors show that context-dependent behavior is related to pupil size, which is related to arousal/neuromodulation, but it would be helpful to have some idea of what neural mechanisms underlying adaptive decision-making are actually impacted by this neuromodulation. Lacking neural data to address this question (e.g., from a brain region proposed to be involved in the accumulation process), at least more discussion of this would be helpful. Essentially, I'm unsure of how to interpret the pupil results: the argument that temporal context affects instantaneous evidence encoding in MT that then

drives the accumulator is very clear, but I am a bit confused about what, mechanistically, I should think about the effect of neuromodulation doing.

We thank the reviewer for this thoughtful comment and agree that the mechanistic interpretation of the pupil results could be made clearer. We acknowledge that we cannot directly identify the neural mechanisms underlying the arousal-related contributions to adaptive evidence accumulation from pupil data alone, given that pupil size is an indirect and imperfect proxy for neural (e.g., LC-NE system) activity. However, we can offer some informed conjecture and have added to the Discussion (L469-482) in an effort to elaborate on possible mechanisms.

Reviewer #3 (Recommendations for the authors):

(1) Abstract could be retooled - does not emphasize the pupillometry/arousal results very much, and they are presented more as a control than an independent result.

We agree and have revised the Abstract accordingly.

(2) Do all neural/pupil analyses use only switch trials? Sometimes the figure captions do specify only switch trials, but not everywhere. It would be helpful to specify either in the Methods or at the beginning of each figure caption that all subplots show switch trial results. Also, if you do always use switch trials, it would be useful to see in the Supplement how the non-switch trial results differ from switch trials. It seems like they may in interesting ways based on the behavioral results (supporting a reset of evidence accumulation on switch but not non-switch trials).

We thank the reviewer for flagging these important points. We have added a justification for switch trials (L186-190) as well as clarification about which trial types were used for which analyses (L246-249) and information about trial types to relevant figure captions. We have also added a new Extended Data figure (Extended Data Fig. 9) examining relationships between neural activity and behavior on non-switch trials. As inferred by the reviewer, behavior on non-switch trials is consistent with the use of information from the adapting stimulus.

(3) In Figure 3C, 5B, etc, when computing firing rate for the test stimulus (50-500 ms), are differently sized windows used to compute the rate for different test stimulus durations (since some will be <500 ms)? Or are only trials where the test stimulus duration is > 500 ms used for this analysis?

We thank the reviewer for raising this point. To clarify, the 50–500 ms window does not reflect a fixed window applicable for all trials. Rather, neural activity from 50 ms after test-stimulus onset through test-stimulus offset was included for each trial, with 500 ms serving as the upper bound for trials with longer durations (> 500 ms). We have clarified this in the Methods (L607-610) to avoid ambiguity.

(4) I think it might be better to be consistent with the time windows used for analysis; specifically, to choose either the 50-500 ms window used in Figures 3, 4, and 5B, or the 200- 400 ms window used for the remaining analyses in Figure 5.

We agree that using the same window for all of the analyses would improve consistency, but not doing so provides advantages that we believe take precedent and now describe in more detail. The broader 50–500 ms window used for Figures 3, 4, and 5B was chosen to characterize MT neural activity over a relatively large a time window, ensuring that every trial contributes to each estimate. Because test-stimulus durations were drawn from a truncated exponential distribution (100–1200 ms), restricting these analyses to the 200–400 ms window would have excluded the substantial proportion of trials with durations <200 ms (but would yield similar figures and conclusions). The narrower window used in subsequent

analyses allows us to focus on the conditions that exhibited the biggest modulations of neural activity when comparing them to behavior.

(5) Similarly, provide justification for using only trials ending 375-600 ms after test stimulus onset for the behavioral correlations. It seems reasonable to choose a subset of test stimulus durations where the monkeys' behavior is greater than chance but less than ceiling, but it would be good to specify this so that it doesn't seem arbitrary.

We agree and have added text to make this important point (L249-251).

<https://doi.org/10.7554/eLife.110685.2.sa0>