# Task Demands Change Motion Information Coding in Area MT of the Macaque

**Hayden Scott**[1,2,3], **Klaus Wimmer**[4], **Tatiana Pasternak**[2,3], **Adam C. Snyder**[1,2,3]

[1] Brain and Cognitive Sciences, University of Rochester, Rochester, 14627; USA
[2] Center for Visual Sciences, University of Rochester, Rochester, 14620; USA
[3] Neuroscience, University of Rochester, Rochester, 14620; USA
[4] Centre de Recerca Matemàtica (CRM), Campus de Bellaterra, Edifici C, 08193 Bellaterra, Barcelona, Spain

‡ Email: adam.snyder@rochester.edu

## Summary

Neurons in the primate Middle Temporal (MT) area signal information about visual motion and work together with the lateral prefrontal cortex (LPFC) to support memory-guided comparisons of visual motion direction. These areas are reciprocally connected and both contain neurons that signal visual motion direction in the strength of their responses. Previously, LPFC was shown to display marked changes in stimulus coding with altered task demands. Since MT and LPFC work together, we sought to determine if MT neurons display similar changes with heightened task demands. We hypothesized that heightened working-memory task demands would improve the task-relevant information and precipitate memory-related signals in MT. Here we show that engagement in a motion direction comparison task altered non-sensory activity and improved stimulus encoding by MT neurons. We found that this improvement in stimulus information transmission was largely due to preferential reduction in trial-to-trial variability within a sub-population of highly direction-selective neurons. We also found that a divisive normalization mechanism accounted for seemingly contradictory effects of task-demands on a heterogeneous population of neurons.

*Keywords*: Neural Variability, Information, Population coding, Working memory, MT, primate, vision

‡ Correspondence: Adam Snyder

## Introduction

The Middle Temporal (MT) area of monkey cerebral cortex is a well-studied extrastriate visual area known to have robust selectivity for visual motion features (Britten et al. 1992; Born and Bradley 2005; Pasternak and Tadin 2020). Motion signals represented by the activity of neurons in this area are important for making memory-guided comparisons about visual motion direction (Lui and Pasternak 2011; Wimmer et al. 2016; Katz et al. 2016). Lesions of this area impair the ability to report whether the motion direction of a stimulus matched a previous example or not (Rudolph and Pasternak 1999; Bisley and Pasternak 2000); whereas electrical stimulation of this area can bias the behavior of an animal as though a particular direction of motion was seen when in fact no direction information had been present (Bisley, Zaksas, and Pasternak 2001; Salzman et al. 1992; Salzman, Britten, and Newsome 1990). These findings indicate a key role for MT in representing behaviorally-relevant visual motion information, and demonstrate that MT has an important role in task-related perceptual decisions. One open question, however, is how task demands impact the neural activity in area MT during sensory (e.g., encoding/decoding of stimulus) and non-sensory (e.g., reward contingencies, decision making, etc.) periods of the task.

Area MT acts in concert with another brain area, the lateral prefrontal cortex (LPFC), in order to do memory-guided comparisons of motion direction. LPFC and MT are reciprocally connected (Ungerleider and Desimone 1986; Barbas 1988; Petrides and Pandya 2006), and both are involved in memory-guided motion comparison tasks (Zaksas and Pasternak 2006). Neurons in LPFC contain direction information during motion discrimination tasks, but this selectivity is largely attenuated when animals do not have to make perceptual decisions, or when that information is not relevant to the task (Hussar and Pasternak 2009; Hussar and Pasternak 2012; Hussar and Pasternak 2013). Previous research has also found reduced trial-to-trial variability in LPFC with increased task demands (Hussar and Pasternak 2010), suggesting LPFC neurons may be recruited to support task performance when needed. Because of the interconnected structural and functional relationship between LPFC and MT, such a change in trial-to-trial LPFC variability with task demands may be reflected in MT dynamics and altered processing of stimulus information, but this hypothesis remains untested.

To investigate this question, we measured how activity in MT covaried with the differing demands of an active task and a passive task that used physically identical stimuli. The active task consisted of two random dot motion stimuli separated by a delay, where subjects (Rhesus macaques; *Macaca mulatta*) had to report whether or not the direction of motion in the first stimulus ("S1") matched the second ("S2"). During the passive task, no perceptual decision was required, but the motion stimuli were the same. Thus, despite practically identical sensory conditions, the task demands during active and passive tasks were different.

We hypothesized that heightened task demands during the active task would be reflected in sharper tuning, reduced trial-to-trial variability, and/or modulation gain, in line with prior findings about selective attention to motion direction (Martinez-Trujillo and Treue 2004; Ponce-Alvarez et al. 2013; Cohen and Newsome 2008; Arandia-Romero et al. 2016). At a population-level, we predicted to find lower noise correlations

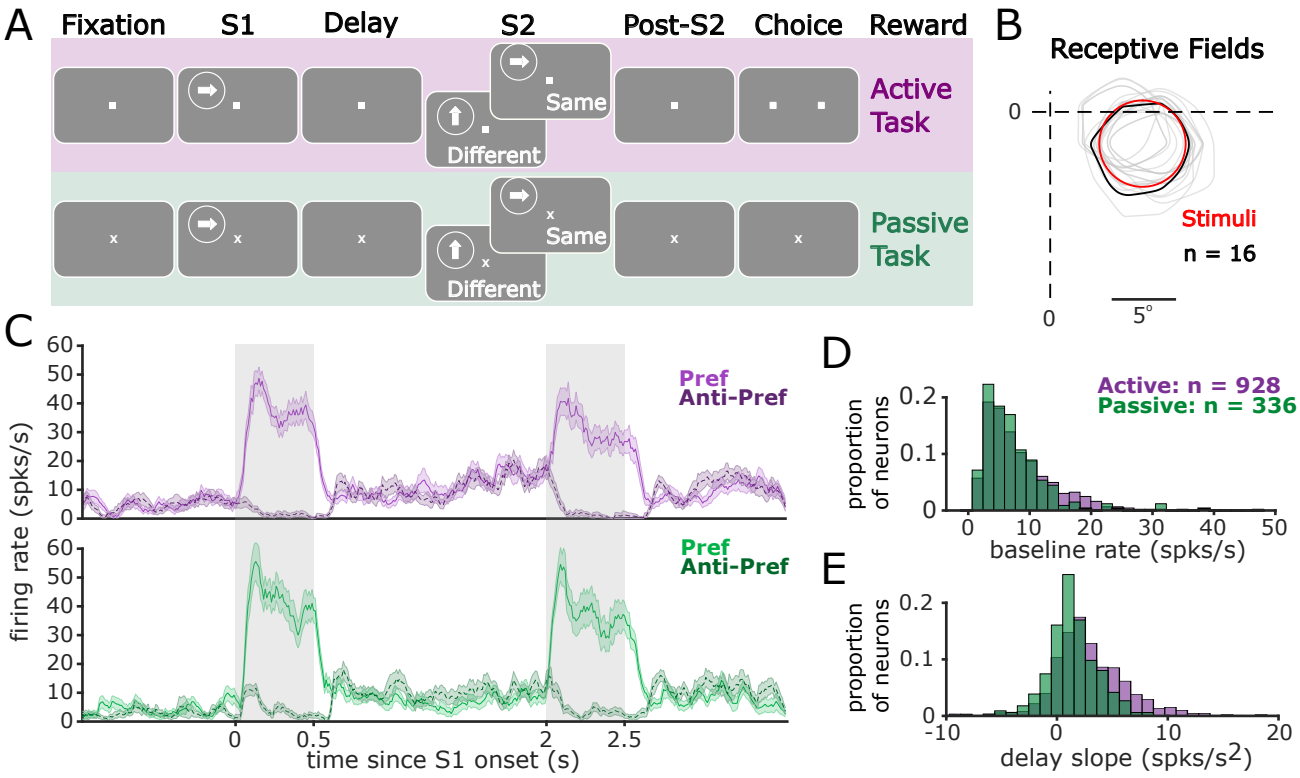


**Figure 1.** *Task design and task modulation of MT firing rates.* A. Trials consisted of a pre-stimulus fixation period, a first stimulus (S1), followed by a 1.5 second delay, then a second stimulus (S2) and a post-test fixation period. After the post-test fixation period subjects were either rewarded (passive task) or had to report a decision with a saccade and rewarded after correct choices (active task). Stimuli moved in one of 8 directions ($0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°$) and the S2 on different trials were always 90° off of S1 (rotated left or right). B. Receptive fields (grey) of each simultaneously-recorded neuron from one example session, along with the location and size of the stimuli (red). Black curve corresponds to example neuron in A. Contours represent isointensity responses at 50% of the peak response. C. PSTHs of an example neuron that was recorded in both tasks. Solid line represents the neuron's response to its preferred motion direction while dashed lines indicate 180 away from preferred ("anti-preferred"). D,E. Histograms of baseline firing rates (D) and firing rate slopes during the delay (E) for all neurons included in the analyses. Slopes were calculated with a linear fit to the neurons' firing rates during a one second window before S2 onset.

in the active task for similarly-tuned neurons, and higher noise correlations for oppositely-tuned neurons, also similar to prior attention studies (Cohen and Maunsell 2009), and in line with hypothesized mechanisms for improved information coding (Averbeck, Latham, and Pouget 2006).

We found that the effects of heightened task demands depended on the amount of task-relevant stimulus information (eg. direction selectivity) neurons conveyed. Specifically, during the active task, neurons carrying strong signals about motion direction showed reduced variability and higher response gain than during the passive task. In contrast, neurons that were visually-responsive, but for which activity signaled direction only weakly, displayed little to no reduction in trial-to-trial variability, as well as a flattening of their tuning curves during the active task (i.e., a *decrement* in direction information when such information was task-relevant). We

found these effects of task demands in MT were consistent with a divisive normalization model (Reynolds and Heeger 2009) in which task demands affected neural populations differently according to their visual feature selectivity.

## Results

Three adult male Rhesus macaques (*Macaca mulatta)* performed delayed visual motion comparison tasks while we recorded neural activity in MT (Figure 1A). In the active task, monkeys indicated whether a second stimulus ("S2") had the same or a 90-degree different direction of motion than a first stimulus ("S1") by making a saccade to one of two choice dots. All three monkeys performed above chance in the active task (m201, 10 sessions: 77.79% ± 1.08% accuracy, m202, 28 sessions: 81.63% ± 0.95%, m317, 16 sessions: 85.77% ± 1.36%; $mean \pm Std$). The passive task always followed the active one, and, except for a different fixation point shape, was identical up to the choice window, at which time no choice target stimuli were shown and no further action was required to receive reward (Figure 1A).

Neurons were recorded with receptive fields of 10.55 ± 0.77 ($mean \pm SEM$ across all neurons from all sessions) degrees of visual angle (DVA) eccentricity. We found that most neurons (1264 out of 1804 recorded neurons: 70.07%) responded to random-dot motion stimuli, as defined by a significant deflection of firing rate from baseline during stimulus windows that lasted at least 50ms, and were included for further analysis. Critically, we observed similar proportions of responsive neurons in both tasks (active: 928 out of 1326, 69.98%; passive: 336 out of 478, 70.29%).

We were interested in the impact of cognitive demands on MT activity. Such

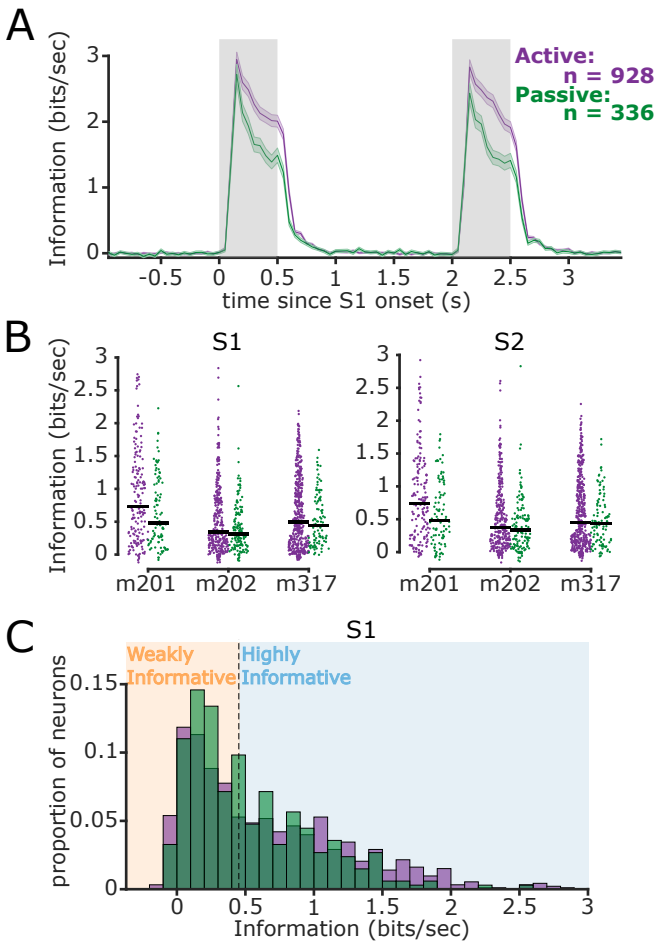


**Figure 2.** *Mutual Information.* A. Average timecourse of mutual information between the firing rate of neurons and motion direction (8 directions), separated by task (active: purple; passive: green). B. Bee swarm plots showing each neuron's mutual information during the stimulus windows separated by monkey. Black lines indicate median values. C. Distribution of mutual information between each neuron's firing rate and motion direction during S1. Neurons below the median (S1: $0.4520\frac{bit}{s}$; S2: $0.4589\frac{bit}{s}$) were considered weakly informative for further analyses, while those higher than the median were deemed highly-informative.

demands varied over the course of a trial in the active task, including periods where stimuli were present (i.e., during S1 and S2) and also periods where stimuli were absent (i.e., pre-S1, the delay, and post-S2). For example, prior to S1, the animal should prepare cognitive resources in anticipation of the stimulus. Then, during S1, the animal must process and store task-relevant information (motion direction). During the delay, this information must then be maintained in working memory and the appearance of S2 may be anticipated. During S2, task-relevant information must again be encoded, the stored information about S1 retrieved, and a comparison judgement made. These dynamic task demands may result in varying effects on MT activity across these different stages. We first describe the effect of task demands on motion-direction signalling by MT activity during stimulus presentations (S1 and S2). Second, we will describe effects of task demands on MT activity when stimuli were absent. Third, we will describe effects of task demands on MT activity during the comparison stage of the task (S2 through decision).

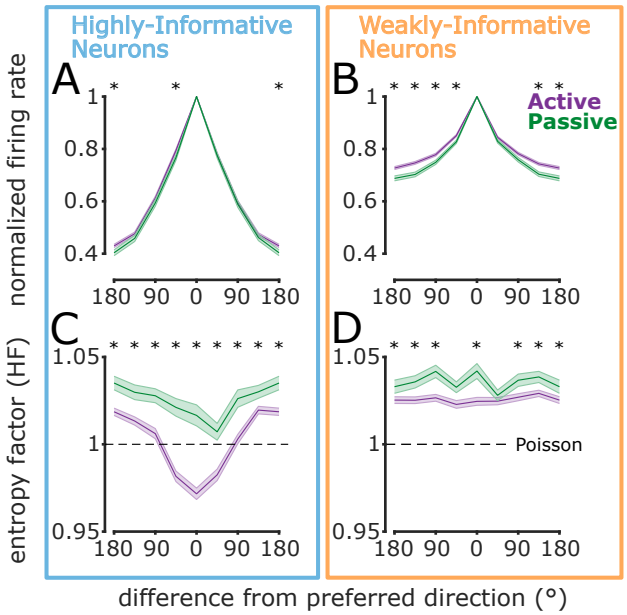


**Figure 3.** *Effects of task demands differed depending on how informative a neuron was about task-relevant information.* A.) Average normalized tuning curves ($mean \pm SEM$ across neurons) for highly-informative neurons in the active (purple) and passive (green) tasks. B.) Same as A, but for weakly-informative neurons. C.) Average trial-to-trial variability ($mean \pm SEM$ across neurons), quantified with entropy factor, for highly-informative neurons. Variability is shown to be a function of motion direction relative to preferred. D.) same as C, for weakly-informative neurons. The difference between S1 and S2 was negligible (tuning: $p = 0.34$; HF: $p = 0.71$ three-way ANOVA between direction, task, and stimulus), so for visualization purposes they have been combined. Asterisks indicate significance at p¡0.05, Kolmogorov-Smirnovtest.

## Task effects in MT depend on neural tuning

We hypothesized that the accuracy of motion direction encoding in MT would improve when that information was task-relevant (i.e., in the active task compared to the passive task). To test this, we measured the mutual information (MI) between neural firing rates and the motion direction of S1 and S2. MI quantifies how much information about one signal is provided by another signal (and vice-versa), and takes into account both the tuning of neural responses with respect to a stimulus as well as trial-to-trial response variability (Hatsopoulos et al. 1998; Quian Quiroga and Panzeri 2009; see Methods). During the active task, MI measured during responses to S1 and S2 was significantly greater than responses to identical stimuli during the passive task (Figure 2A); S1: $D = 0.1196$, $p = 0.0015$, S2: $D = 0.1336$, $p = 2.590 \cdot 10^{-4}$ Kolmogorov-Smirnov test). This indicates that, overall, MT neurons encode more information about motion direction when it is task-relevant.

We hypothesized that changing task

demands might be especially impactful on the activity of neurons that signal task-relevant information. That is, switching from passive viewing to a task requiring judgment of motion direction might have a particularly strong effect on neurons strongly signaling motion direction compared to neurons that did not signal that information so strongly. To test this hypothesis, we split our sample of neurons into two groups, based on whether their MI fell above or below the median information across all neurons, separately calculated for a 500 ms time window following the onset of each stimulus (S1 median: MI $= 0.4520\frac{bit}{s}$; S2 median: MI $= 0.4589\frac{bit}{s}$). We call neurons with MI above the median, "highly-informative", and we call neurons with MI below the median "weakly-informative". Note that MI was unimodally distributed, so that "highly-" and "weakly-informative" labels should be taken as relative, rather than categorical. We tested several MI values as thresholds for dichotomizing our sample and found that the qualitative pattern of results and our conclusions were highly robust to substantial variation of the precise threshold chosen. Thus, we ultimately chose the median as the threshold for dichotomization because that results in two equally-sized groups of neurons, which is advantageous for comparison purposes. After splitting the neurons into these two groups, highly-informative neurons still had significantly more information in the active task (S1: $1.0951 \pm 0.0220\frac{bit}{sec}$, S2: $1.0835 \pm 0.0229\frac{bit}{sec}; mean \pm SEM$) compared to the passive task (S1: $0.9187 \pm 0.0305\frac{bit}{sec}$, S2: $0.8842 \pm 0.0330\frac{bit}{sec}; mean \pm SEM$ Kolmogorov-Smirnov: S1: $D = 0.1717$, $p = 0.0019$; S2: $D = 0.1989$, $p = 1.804 \cdot 10^{-4}$). Interestingly, we found a significant reduction in MI during S1 for weakly-informative neurons in the active task (S1: $D = 0.1425$, $p = 0.0087$; Kolmogorov-Smirnov test); though this reduction did not reach significance during S2 (S2: $D = 0.0934$, $p = 0.1959$; Kolmogorov-Smirnov test). These neurons signal less task-relevant information during the active task, relative to passive. These results suggest that with heightened task demands, task-relevant information becomes disproportionately represented in a smaller number of highly-informative neurons.

## Heightened Task Demands Flatten the Tuning Curves of Less-relevant Neurons

We found that mutual information between neuron firing rates and motion direction changed when motion direction was task-relevant (active task) compared to when it was not (passive task). Since MI takes into account both stimulus tuning and trial-to-trial variability, the observed changes in MI that could be consistent with a change in tuning, a change in trial-to-trial variability, or a mixture of both. To better characterize the nature of the task-demand effects we found, we separately analyzed tuning and trial-to-trial variability.

We tested for differences in tuning for each neuron by calculating the average firing rate for each direction condition and normalizing by the response to the preferred direction. The main effect of 'stimulus' (S1 vs S2) was negligible (for tuning analysis: $p = 0.34$; for HF analysis: $p = 0.71$; three-way ANOVA, see methods), so for visualization purposes we combined them. We found little difference in the average tuning profiles of highly-informative neurons between the active and passive tasks (Figure 3A). In contrast, we found a significant flattening of tuning curves for weakly-informative neurons in the active task compared to the passive task

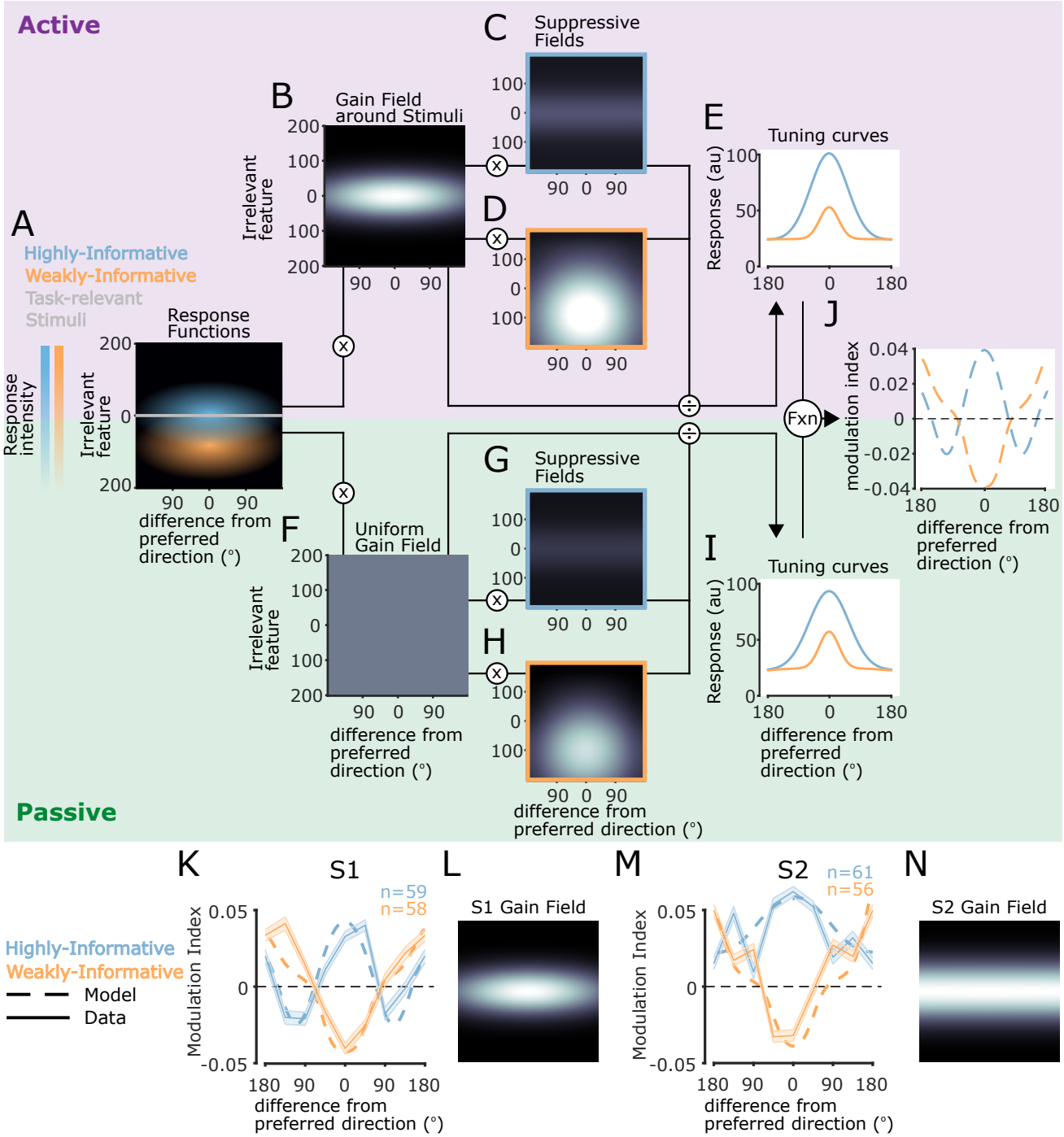


**Figure 4.** *A Divisive normalization model explained the pattern of modulation indices for both highly and weakly-informative neurons* A.) The response functions depicted in a 2D feature space. White line represents the stimuli shown in this experiment. Highly-informative neurons (blue) are centered over these stimuli, whereas weakly-informative neurons (orange) had their preferences centered at a different part of an unobserved feature axis. B.) The gain field for the active task. It is centered on an enhancement around the relevant stimuli for the task. The excitation field was multiplied by different gain fields for S1 (Panel B) and S2 (Panel N). C.) the suppressive field for highly-informative neurons. This is the normalization step where neurons' activity affects others in the population. D.) Same as C but for weakly-informative neurons. E.) The resulting tuning curves from the model for two neurons at the center of their respective population distributions. F-I.) Same as B-E, but for the passive task. J.) The resulting modulation indices for the model neurons' tuning curves in E and I. (see Methods). K.) Modulation indices for matched neurons (solid line), as well as model fit (dashed line) for S1. L.) Model-derived gain field during S1. M-N.) same as K-L but for S2.

(Figure 3B). In weakly-informative neurons, the normalized response to the anti-preferred (180° away from preferred) motion direction was significantly higher for neurons in the active task (Kolmogorov-Smirnov test: S1: $D = 0.1412$, $p = 0.0096$; S2: $D = 0.1318$, $p = 0.0197$). Significant increases in normalized firing rates were also found for intermediate directions between preferred and anti-preferred during S1 ($-90°$: $D = 0.1535$, $p = 0.0037$; $-45°$: $D = 0.1460$, $p = 0.0066$), and S2 ( $-90°$: $D = 0.1356$, $p = 0.0151$; $45°$: $D = 0.0669$, $p = 0.5916$). This finding was further supported by the pattern of modulation indices (see Methods)) for this weakly-informative group of neurons (Figure 4K,M), which showed a suppression of response to preferred motion and enhancement of anti-preferred directions. This means that during the active task, neurons that weakly encoded direction had flatter (broader) tuning profiles, which in isolation would constitute an unexpected deficit to direction coding.

## Increased Task-demands Decrease Trial-to-Trial Response Variability

Since we found a substantial improvement in mutual information for the active task that was not well accounted for by changes in tuning, we reasoned that the improved encoding quality may be due to reduced variability of neural responses. To test this hypothesis, we calculated the entropy of neural responses for each task and normalized measurements relative to those of a simulated Poisson process matched for intensity to the recorded neurons (entropy factor (HF); see Methods).

We found an overall reduction in HF in the active task, relative to the passive task(Figure 3C,D). This reduction in trial-to-trial variability was consistent for both highly-informative and weakly-informative neurons. For highly-informative neurons, we found a reduction in trial-to-trial variability that was a function of motion direction (Figure 3C). These highly-informative neurons in the active task displayed a large reduction in trial-to-trial variability during their preferred stimuli, with weaker effects as motion direction diverged from this angle. Weakly informative neurons showed no such preferential variability reduction, but did show a modest reduction in variability, independent of direction condition (Figure 3D).

To summarize, the reduction in variability we found was consistent with the robust overall improvement in MI that we found. Compared to the weak effects on tuning (Figure 3A,B), the strong and consistent reductions in trial-to-trial variability lead us to conclude that effect is likely a primary contributor to the higher information content during the active task.

## Task Demands Alter the Noise Correlations of Highly-Informative Pairs of Neurons

We found the average information about motion direction of individual MT neurons tended to increase when that information was task-relevant. However, the information contained in a population can be less than the sum of its individual neurons depending on the "noise correlation" of those neurons (Averbeck, Latham, and Pouget 2006). To test whether task demands also affected the noise correlation of MT neurons, we calculated noise correlations for pairs of simultaneously-recorded neurons across repeated presentations of each motion direction condition. Previous studies of noise correlations in visual populations have found that neurons with similar stimulus preferences tend to have greater noise correlations (Kohn and Smith 2005). To ac-

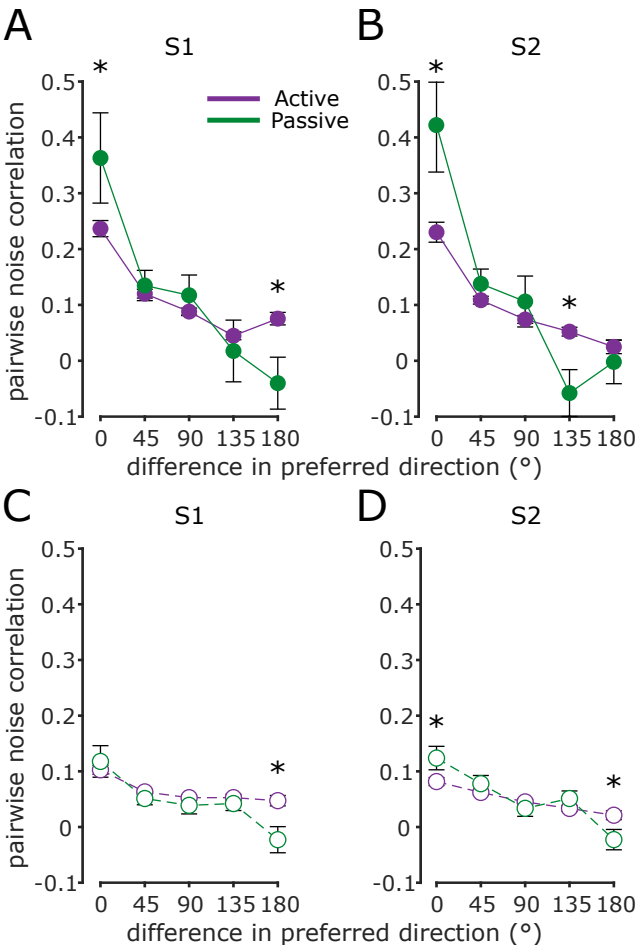


**Figure 5.** *The effect of task engagement on correlated variability* A.) Pairwise spike-count correlations during S1 between highly-informative neurons in the active and passive tasks, separated by the angular difference between their preferred motion direction. B.) Same as A for S2. C.) Pairwise correlations for "mixed pairs" of neurons during S1. Mixed pairs consisted of one highly-informative and one weakly-informative neuron. D.) Same as C for S2.

count for this, we grouped pairs of neurons according to the difference in their preferred motion direction. We found that noise correlation was inversely related to the similarity of preferences between the neurons in the pair (Figure 5), as has been observed for other brain areas (Smith and Kohn 2008; Smith and Sommer 2013; Khanna, Snyder, and Smith 2019). We also found that pairs of neurons in the active task (compared to pairs in the passive task) had weaker noise correlations when they were similarly tuned, and stronger noise correlations when oppositely tuned, which is consistent with theoretical coding improvements of a population (Averbeck, Latham, and Pouget 2006; Ruff and Cohen 2016). The effect of task demands on noise correlations was almost entirely limited to pairs of highly-informative neurons (S1 $0°$ : $p = 0.0052$; $180°$ : $p = 4.44 \cdot 10^{-4}$; S2 $0°$ : $p = 1.97 \cdot 10^{-4}$; $135°$ : $p = 3.52 \cdot 10^{-5}$, Two-sample t-test). Noise correlations between one highly-informative and one weakly-informative neuron ("mixed pairs") did not depend on task demands, with the exception of neurons with opposite direction preferences (S1 $180°$ : $p = 0.0017$; S2 $0°$ : $p = 0.0171$; $180°$ : $p = 0.013$, two-sample t-test).

## A Divisive Normalization Model Accounts for Heterogenous Effects of Gain Modulation Across the Population

We found that MT neurons that were highly informative about motion direction were even more informative in the active task compared to the passive task, while MT neurons that were weakly informative showed the converse pattern (Figure 2A ;Figure 4K,M). That any neurons became less informative about motion direction when motion direction became task relevant was somewhat surprising, as removal of task-relevant information would seem at first blush sub-optimal. To better understand this phenomenon, we considered that neurons in MT have mixed selectivity for multiple features (Maunsell and Van Essen 1983; DeAngelis and Uka 2003; Rodman and Albright 1987; Kim, Angelaki, and DeAngelis 2015). Two neurons may be tuned similarly for motion direction, but prefer different motion speeds, for example. This is illustrated in Figure 4A, where one type of neuron (blue)

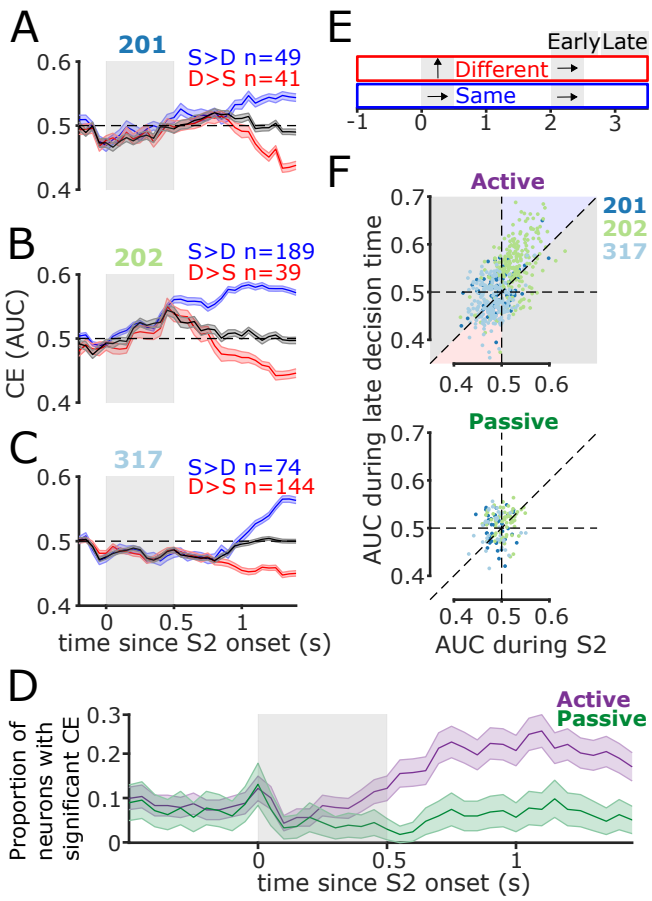


**Figure 6.** *Comparison Effects increase from stimulus to decision* A. Time course of comparison effects (area under the receiver operating characteristics curve: AUC) from S2 onset through end of trial for m201. Neurons were split into three groups: significantly same preferring ($S > D$, blue), different preferring ($D > S$, red), and not significant (black). B-C. Same as A for the other two monkeys. D. Time windows for computing comparison effects in panels E and F. E. Comparison effects from late decision time vs S2 for the active task. Dot color represents individual subject. Values greater than 0.5 indicate $S > D$ while values lower than 0.5 indicate $D > S$. Blue area indicates neurons that had same-preferring effects throughout both windows, but whose same-pref signal increased over time. Red area indicates the same for different preferring neurons. Grey area represents a dynamic zone, where neurons either flipped sign across the windows or had decreasing effects over time. F. Same as E for the passive version of the task.

prefers the stimuli we used in the experiment (white line), while the other type of neuron (orange) prefers a slightly different value along another feature dimension (eg. speed, spatial frequency, disparity, etc). Thus, while the information signalled about direction is similar for the two neuron types, one type (blue) is more task-relevant than the other (orange), due to stimulus parameters that were untested in the experiment.

In this scenario, we hypothesized that a gain applied to neurons whose stimulus preferences over multidimensional feature space were most-relevant for the task might result in a traditional gain effect, which results in a suppression of neurons whose preferences were "nearby" in feature-space via divisive normalization. Divisive normalization computes the ratio of the response of an individual neuron to the summed activity of a pool of neurons; it can be implemented through suppressive interactions between neurons, and has been proposed as a canonical neural computation (Carandini and Heeger 2012). In our case, because we did not measure tuning across stimulus features other than motion direction, we tested the tenability of this hypothesis with an adaptation of the divisive normalization model of Reynolds and Heeger (2009) (Figure 4). That model consists of three main components (1) an excitation field that describes the response of a neuron to stimulus input (Figure 4A), (2) a gain field that describes an excitatory modulatory influence (Figure 4B,F), and (3) a suppressive field that describes a suppressive influence between neurons (Figure 4C,D,G,H). We adapted the model to incorporate two heterogeneous neural populations with overlapping gain fields (Figure 4B,F) and suppressive fields (Figure 4C,D,G,H). Weakly-informative neurons were specified to have the same preferred direction as highly-informative

neurons, but a different preference along a second feature dimension.

We modeled stimulus responses for these two groups of neurons in a passive condition where the gain field was assumed to be uniform (Figure 4F), and an active condition, where the gain field was free to be non-uniform (Figure 4B). We could then calculate modulation indices for the modeled neurons (Figure 4J), just as we did for real MT neurons (Figure 4K,L). In this model, we used simplex search (see Methods) to optimize the center and width of the excitation fields (neural tuning), gain field, suppression fields, and baseline terms to best fit the average modulation indices of neurons recorded in both tasks. The excitation field was fixed across active and passive, but the gain and suppressive fields were free to vary. Highly-informative and weakly-informative model neurons shared a gain field, but were allowed to have different suppression fields to reflect differences in their contribution to the normalization step.

A simple prediction of this model is that on average, weakly-informative neurons should respond to their preferred motion direction with lower firing rates than highly-informative neurons, which was true ($p = 3.7 \cdot 10^{-62}$ two-sample t-test on average response to preferred during S1). Additionally, the model explains the inverted-gain pattern seen in the weakly-informative neurons (Figure 4K,M) through a suppression mechanism from the highly-informative neurons. The model also indicated that a more narrow attention field during S1 (Figure 4L,N) could account for suppression of the highly-informative neurons responses to 90° off preferred, which may result in a finer estimate of motion direction. Taken together, our model suggests that task demands modulate neurons proportional to their tuning with regard to the perceptual task at hand.

## Task demands modulate non-stimulus firing rates in MT

We next turned our attention to the effects of varying task demands on MT activity during periods when dot motion stimuli were absent (pre-S1, during the delay between S1 and S2, and post-S2). Because no stimuli were present, effects during these intervals would be due only to cognitive processes (e.g., anticipation of S1 or maintenance of memoranda during the delay).

During the pre-S1 fixation period, we found a significant increase in the average firing rate in the active version of the task compared to the passive version of the task (Figure 1C: Kolmogorov-Smirnov test: $D = 0.095$; $p = 0.022$), suggesting that increased task demands enhance excitability of MT neurons even in the absence of visual stimulation. The pre-S1 period was highly predictable in this task (1000ms post fixation), so we expected to see a gradual increase ("ramping") of firing rate during this period, reflecting anticipation. We found significant pre-S1 ramping, which we quantified as the slope of a linear fit to FR vs. time during a one second window before S1 onset ($1.740 \pm 0.0024 \frac{spikes}{s^2}$). Next we considered the delay period between S1 and S2. Because the timing of this delay was also highly predictable in our task (always 1.5 seconds), animals could anticipate the appearance of S2. We also found significant ramping during the delay period ( $2.5657 \pm 0.003 \frac{spikes}{s^2}$ $mean \pm SEM$), where, on average, neurons in the active task had higher slopes than neurons in the passive task (Figure 1E: Kolmogorov-Smirnov test: $D = 0.248$; $p = 8.425 \cdot 10^{-14}$). This increased ramping in the active task suggests

an effect of anticipating the S2 stimulus, and was stronger than the pre-S1 ramping (Two-sample F test: $F = 1.364$; $p = 3.716 \cdot 10^{-8}$ $N = 1264$), suggesting either more intense or more reliable anticipation, or a further role of ramping during the delay than just stimulus anticipation, likely reflecting additional task demands related to the preparation of the S1 and S2 comparison.

We also asked whether MT activity might continue to be selective for the motion direction of S1 during the delay, which could be consistent with a working-memory-related maintenance process. However, we found that MI between MT activity and S1 motion direction rapidly fell to levels consistent with the null hypothesis after S1 offset, and this did not differ based on task demands (Figure 2A).

## MT Activity is Affected by Cognitive Aspects of Memory-guided Visual comparisons

During S2 of a given motion direction, our active task required subjects to make a saccade to the right if the direction of the preceding S1 was the same ("same trial"), and a saccade to the left if the direction of the preceding S1 was different ("different trial"). We asked if MT neurons' responses to physically identical S2 stimuli differed whether they came from "same" or "different" trials, which we termed a "comparison effect". A greater response to S2 on "different" trials compared to same trials ("$D > S$") could be consistent with a response adaptation effect, while a greater response to S2 on "same" trials ("$S > D$") would be more consistent with a facilitation effect (Kohn and Movshon 2003). We would expect that if adaptation occurs, we should find similar effects in the active and passive tasks, and any task effect should indicate a greater role for memory-guided comparison processes.

We calculated comparison effects as the area under the receiver operating characteristics curve (AUC) between identical stimuli (S2), during same and different trials (Figure 6E). We found that some neurons responded more robustly to the same stimulus during S2 when the direction matched that of S1, so called "same-preferring" neurons ($S > D$ ; m201: $n = 49, 29.2\%$; m202: $n = 189, 59.6\%$; m317: $n = 74, 16.7\%$). We also found "different-preferring" neurons, in that they responded stronger to stimuli that differed from S1 ($D > S$ ;m201: $n = 41, 24.4\%$; m202: $n = 39, 12.3\%$; m317: $n = 144, 32.5\%$).

We noted that the three subjects we tested had significantly different proportions of same-preferring and different-preferring neurons ($\chi^2 = 154.324, p = 3.083 \cdot 10^{-34}$), and we wondered whether this might be related to animals' idiosyncratic strategy for performing the task. For example, a "different" decision might be more automatic than a "same" decision for a particular animal, and be associated with less LPFC involvement (Dias et al. 2006). In this case, we would predict that the animal would be biased toward making "different" reports, and would have a smaller relative proportion of different-preferring neurons. In line with this notion, we found that subjects with more $S > D$ neurons (m201, m202) had significant biases for choosing "different" responses over "same" responses (m201: 2129 same choices vs 2303 different choices, $p = 0.0019$, two-tailed binomial test; m202: 3683 same choices vs 3971 different choices, $p = 7.045 \cdot 10^{-5}$). The third subject had more $D > S$ neurons, and a trend for "same" choice bias (m317: 2750 same choices vs 2664 different choices, $p = 0.2113$; two-tailed Binomial test). This result suggests subjects may use LPFC activity to counter their intrinsic choice biases.

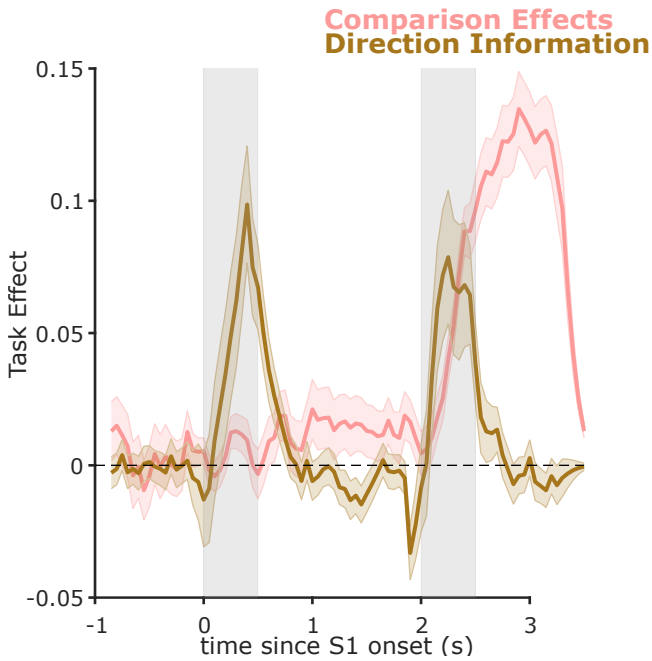


**Figure 7.** *Relative sensory and cognitive effects of heightened task demands.* Proportion of neurons with significant effects in active, minus passive, for MI and comparison effects. We found about 10-15% more neurons are recruited in area MT for both sensory and cognitive aspects of the task compared to passive fixation.

### Heightened Task Demands Recruit Similar Numbers of MT Neurons Across Multiple Phases of the Task

Since MT neurons signal both sensory (MI, Figure 2A) and cognitive (anticipatory ramping, Figure 1E; comparison effects, Figure 6) components of the task, we next sought to compare the relative extent to which task demands impacted MT's involvement in those processes. Because each effect (MI, ramping, comparison) was measured in different units, a direct comparison of task effect sizes was not possible. Instead, we compared the relative proportion of MT neurons for which each effect was significant across the two tasks. In the active task, 815 out of 928 neurons had significant pre-S2 ramping (87.8%), with only 76.5% in passive (257 out of 336 neurons). This constitutes a significant increase in the proportion of neurons with ramping effects across task ($\chi^2 = 24.60, p = 7.04 \cdot 10^{-7}$). We found a similar effect size for both mutual information (MI) and comparison effects (CE) during the active and passive tasks, with about 10-15% more neurons recruited in the active task, relative to passive (Figure 7).

## Discussion

We hypothesized that neurons in the primate Middle Temporal (MT) area behave differently when task demands increase in order to improve information processing for goal-oriented behavior. We tested this hypothesis by recording activity of MT neurons in monkeys performing two tasks, one active and one passive. We found that in the active compared to the passive task, (1) MT activity carried more information about motion direction during visual stimulation, (2) firing rate ramping was increased during non-stimulus periods, and (3) comparison effects were strengthened.

### Varying Task Demands Alter the Amount of Direction information Contained in MT Activity During Visual Stimulation

We found that firing rates of neurons in MT contained more direction information during the stimulus periods in the active task than in the passive task (Figure 2A). We then asked whether the magnitude of this improvement was related to the alignment of the neurons' tuning and the stimulus parameters used. We reasoned that cognitive factors might preferentially target neurons that were highly-informative about task-relevant information, compared to neurons that were only weakly-informative. In this scenario, the effect of increasing task demands would be predicted to

be greater for the highly-informative neurons, and may be reflected in differential changes to tuning or variability. However, the extent to which tuning and variability are each influenced by cognitive factors has been a subject of intense research (Renart and Machens 2014; Gilbert and Li 2013), and each has different implications for computation and neural mechanism. To better characterize the role of task demands on motion direction signaling in MT in the context of this prior work, we separately examined tuning and variability. The average, normalized tuning profile for highly-informative neurons' tuning curves was remarkably unchanged by our manipulation (Figure 3A). However, inspection of neurons that we were confident were recorded in both task contexts ("matched" neurons) revealed a subtler pattern of modulations consistent with a change in tuning (Figure 4K,M). In contrast, we found that weakly-informative neurons became even *less* informative about motion direction when that information was more task-relevant due to a flattening of tuning curves (Figure 3B; Figure 4K,M). As a result, these neurons became less tuned for motion direction when that information was more relevant.

From a normative standpoint, an ideal observer could use the information signaled by these neurons, albeit weakly, to improve decoding of motion direction from the activity of MT populations. However, this normative model assumes that the brain can utilize information encoded in neural activity in arbitrary ways, which may not be true in practice. Instead, the brain may preferentially use only the most informative neurons, in which case the activity of less informative neurons may be suppressed via divisive normalization (Carandini and Heeger 2012).

To test this notion, we adapted the normalization model of Reynolds and Heeger (2009). We found that this model captures the effects of task demands on both neural populations, as well as differences in modulation between S1 and S2. For neurons matched across task, we found a traditional gain-like pattern for highly-informative neurons, but a converse-gain pattern for weakly-informative neurons. In addition to the two populations of neurons, we saw a different pattern of modulations between S1 and S2, which was explained through a wider attention field during S2 along the relevant feature dimension. We believe that this may reflect different strategies involved in processing the sensory information for S1 and S2. An accurate estimate of motion direction was more important during S1, as small uncertainty about direction would lead to a much more difficult time correctly doing the trial (whether S2 was the same or different). S2 on the other hand, only required a rough estimate of motion direction to do the task (being off by 45° would not impact the decision). This is because "different" trials were always 90° away from the direction during S1, so a more precise representation of S2 motion direction was not required. Taken together, these results suggest that the improved direction encoding displayed by our sub-population of highly-informative neurons suppresses neurons with less-task-relevant feature tuning. This model accounts for the potential role of tuning in task-demand-related changes in mutual information (MI), but does not take into consideration another equally-important component, trial-to-trial response variability, which we examined next.

By measuring entropy factor (HF), we found that MT neurons in the active version of the task were less variable in their response to visual stimuli (Figure 3C,D). We also found a strong relationship between direction prefer-

ence and variability for highly-informative neurons. Those neurons had the largest reduction in trial-to-trial variability during presentations of their preferred stimulus, with weaker effects as direction diverged from preferred. This cannot be due to the increased firing rate for those directions, as HF accounts for rate effects (see section Methods), and a rate confound would result in the opposite trend (that is, typically variability is biased downward for lower rates, but we found higher variability at lower rates; Churchland et al. 2010). The reduction in trial-to-trial stimulus response variability we observed when task demands were elevated is reminiscent of the finding of Mitchell, Sundberg, and Reynolds (2007), who found general reductions in Fano Factor in V4 responses when animals selectively attended to a stimulus compared to when that stimulus was not attended. The novel and more surprising result we found is that the strength of variability reduction for highly-informative neurons depended on whether the stimulus was preferred (great reduction in variability) or less-preferred (less reduction in variability). This suggests that one effect of elevating task demands may be a preferential improvement to information transmission reliability according to neurons' feature tuning.

In addition to changes in individual-neuron response variability, we hypothesized that task-demands would have a marked impact on neural covariability. It is well-appreciated that the structure of correlated trial-to-trial variability across a population impacts information coding (Panzeri et al. 1999; Moreno-Bote et al. 2014; Bartolo et al. 2020), so we calculated spike-count noise correlations between pairs of simultaneously-recorded neurons. We found that neurons with similar preferred directions had lower noise correlations in the active task, while ones with opposite preferences were higher (Figure 5A,B), consistent with earlier studies of attention (Cohen and Maunsell 2009; Ruff and Cohen 2016; Smith and Kohn 2008). A future study could test the benefits of this observation more directly by attempting a neural population decoding approach (Semedo et al. 2019; Snyder, Yu, and Smith 2018; Kaufman et al. 2015; Quian Quiroga and Panzeri 2009). Interestingly, the effect of heightened task demands on noise correlations was largely limited to pairs of neurons that were both highly-informative. So called "mixed" pairs (one highly-informative and one weakly-informative) of neurons displayed a much weaker task effect (Figure 5C,D). This is consistent with our finding at the single-neuron level that elevating task demands leads to preferential improvement of MI for neurons that were highly-informative for the task.

## Effects of Task Demands on MT Activity during Non-sensory Intervals

We found significant changes in non-stimulus firing rate (FR) in the active task compared to the passive task, such as higher baseline FR, as well as larger FR ramping during the late delay period (Figure 1C). Ramping of neural firing rates has traditionally been explored in the context of motor preparation (Ding 2015; Narayanan 2016), and decision making (Shadlen and Newsome 2001), where it has been typically linked to anticipation. Finding anticipatory increases in firing rate for sensory neurons in MT is interesting, and is not likely explained through either decision making or motor planning, as neither a correct decision or appropriate motor plan can be made before first seeing S2. More likely, the ramping we observed is related to preparing cognitive and attentional resources for the perceptual phases

of the task, which had highly predictable timing (Hussar and Pasternak 2010; Bisley et al. 2004). This line of thought has been explored in the context of another brain region, V4 (Snyder, Yu, and Smith 2018; Luck et al. 1997), where anticipatory ramping has been linked to interactions between V4 and prefrontal cortex (Snyder, Yu, and Smith 2021). Taken together, these results suggest anticipatory signals may originate in prefrontal cortex and influence visual cortex via top-down feedback.

## Effects of Task Demands on MT Comparison Effects

In addition to the more strictly sensory aspects of the task, there is a comparison component which requires additional cognitive resources such as working memory. We found significantly more neurons conveyed comparison signals in the active task. It is possible that adaptation (a weaker response to repeated stimuli) may partially explain the $D > S$ neurons (Kohn and Movshon 2003). However, this is unlikely to fully explain our results because many neurons had $D > S$ signals that strengthened over time, well outside of the stimulus window (Figure 6D), which would be inconsistent with adaptation. Interestingly, the prevalence of significant CE neurons was a counter-indication of the behavioral bias for two of the three subjects, with a trend on the third. For example m202 had several times as many $S > D$ neurons, but had a significant bias for making "different" choices. It may be the case that each subject has a default choice requiring less prefrontal activity to execute, and requires prefrontal activity to counteract that default choice. As in, m202 naturally defaults to choosing "different" and so requires more $S > D$ neurons to accumulate "same" evidence to counteract this prepotent response. This idea follows naturally from previous work that demonstrates a strong correlation between CE and choice probability (Lui and Pasternak 2011) in area MT. That is, we found more MT neurons preferring 'same' trials in animals that had a greater tendency to make 'different' choices. This may at first blush seems inconsistent with our prior report that neurons preferring 'same' trials also show choice probability signals predicting 'same' choices. However, an important distinction is that the current finding does not concern trial-by-trial prediction of behavior, but rather overall behavioral biases. That is, the two reports are not inconsistent because an animal biased to make 'different' choices may have a greater prevelance of same-preferring neurons that predict choices trial-by-trial. Together, this work ties nicely with previous studies in both monkeys and humans, that demonstrate a strong relationship between lateral prefrontal cortex (LPFC) activity and behavioral inhibition during cognitively-demanding tasks (Dias et al. 2006; Durston et al. 2002).

## Conclusion

In the present study, we investigated the effects of task demands on information processing in area MT of the macaque. We found substantial differences in variability and tuning profiles depending on the degree to which a neuron represented task-relevant information in its firing rate. These differences can be accounted for with a divisive normalization model that accounts for neural heterogeneity. Future work should investigate more closely how tuning in high-dimensional feature spaces changes with altered task demands by carefully manipulat-

ing stimuli along multiple feature axes. In addition to these sensory components of the task, MT neurons also show stronger cognitive signalling with heightened task demands, such as anticipatory ramping and heightened comparison effects. These results highlight the role of MT populations in all phases of memory-guided comparisons of visual motion direction, and, combined with the previous findings in LPFC, suggest a continual interplay between prefrontal and extrastriate visual cortical areas during cognitive task performance.

## Acknowledgements

This work was supported by NIH grant R01EY011749 (T.P. and A.C.S.); NIH grant R00EY025768 (A.C.S.); a Sloan Research Fellowship (A.C.S.), NARSAD Young Investigator Award (A.C.S.), as well as the Spanish State Research Agency together with the European Regional Development Fund (grants RYC-2015-17236 (K.W.), BFU2017-86026-R (K.W.), PID2020-112838RB-I00 (K.W), and, through the Severo Ochoa and María de Maeztu Program for Centers and Units of Excellence in RD, CEX2020-001084-M). KW thanks CERCA Programme/Generalitat de Catalunya for institutional support.

## Author Contributions

This project was conceptualized by T.P. and K.W. Data acquisition was done by T.P. Data analysis was done by H.S., with guidance from A.S. and K.W. Original draft of manuscript was written by H.S. with A.S., K.W., and T.P. providing editorial and revision feedback.

## Declaration of Interests

The authors declare no competing interests.

## Methods

### Resource Availability

*Lead Contact* Requests for further information should be directed to the lead contact Adam Snyder (adam.snyder@rochester.edu).

*Materials Availability* This study did not generate any new reagents.

*Data and Code Availability* Data and code used for this paper are made available by request through the lead contact.

### Subjects

Subjects used in this study were three adult male rhesus macaques ages 7, 8, and 7 years, for m201, m202, and m317. All training, surgery, and experimental procedures were performed in accordance with the National Institutes of Health *Guide for the Care and Use of Laboratory Animals* and were approved by the University of Rochester Committee for Animal Research. Surgery was performed using aseptic technique with isofluorane general anesthesia and perioperative opiate analgesics. An initial surgery was performed to implant a surgical steel head holder restraint embedded in bone cement secured to the cranium with ceramic bone screws. After behavioral training, a subsequent surgery was performed to implant a PEAK canula (19 mm diameter; Crist Instruments, Hagerstown, MD) over MT.

### Data Acquisition

Recording chambers enclosed the craniotomies and contained 1mm-spaced CILUX grids (Crist Instruments). Thirty-two-channel S-probes (Plexon, Dallas Texas) were lowered through custom-made steel guide tubes, which

were themselves inserted just low enough to penetrate the dura. We then drove the electrodes using a NAN electrode drive (NAN Instruments, Nof Hagalil, Israel). Recordings were done by coordinating a Plexon (Dallas, Texas) Multichannel Acquisition Processor and the data acquisition system TEMPO (Reflective Computing).

### Receptive Field Mapping

Prior to each recording session, receptive fields were mapped in order to identify where the stimuli should be displayed. The initial manual phase consisted of moving a patch of dot motion around on the stimulus display with a computer mouse while listening to spiking activity via an audio monitor. Afterwards, we presented small dot motion patches (1°) in different directions on the vertices of a grid spanning the likely RF area identified in the manual step. Stimuli for each session were placed such that they covered the aggregate receptive field area for the recording (Figure 1B).

### Stimulus Presentation

Stimuli were presented at 75 Hz monitor refresh rate on a 19-inch IIyama Vision Master Pro-513 monitor with 1,152 by 870 pixel resolution. They were 100% motion-coherence random dot kinematograms in one of 8 directions $(0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°)$ placed in a circular aperture fit to the size of the receptive fields, with a dot diameter of 0.03° and a density of $3\frac{dots}{deg^2}$. Dot luminance was $15\frac{cd}{m^2}$, shown on a dark background of $0.1\frac{cd}{m^2}$. Monkeys were required to maintain fixation within 2° of a central dot throughout the trial until the response interval, and eye position was monitored with an ISCAN infrared eye-tracking device.

### Task Design

Trials started once subjects fixated on a central dot. They then had to maintain fixation for 1 second before S1 onset. Both S1 and S2 lasted 500ms. S1 was followed by a 1.5 second delay period, and then the second stimulus S2 was presented. Subjects then had to wait an additional 500ms after S2 offset before either a choice was required (active task), or they were simply rewarded (passive task). For the active task, the central fixation point was extinguished at the beginning of the response interval, and two identical choice targets were presented 5 degrees to the right and left of fixation, and the animal indicated its judgement as to whether S1 and S2 were the same direction or not by making a saccade to the right or left choice target, respectively. In most recording sessions, we collected data for both the active and the passive tasks. We always collected data for the active task before the passive task, because animals would have been unmotivated to perform the more difficult active task if they had performed the easier passive task first.

### Data Analysis

Spike sorting was done offline with Plexon sorter (Version 3.3.2) by using principal component analysis (PCA) on the spike waveforms and manually clustering. A neuron was included for analyses if there were at least ten trials in each direction condition, and it was visually responsive as defined by a significant deflection from baseline firing rate that persisted at least 50ms during stimulus presentation (two-sample t-test, $\alpha < 0.05$). A total of 1264 neurons across both tasks passed these criteria (m201: n = 263, m202: n = 454, m317: n = 547), with 928 neurons recorded in the active task and 336 neurons recorded

in the passive task (see "Modulation Index" below for matching criteria). All subsequent analyses were done using custom routines for Matlab 2020a (The MathWorks, Inc, Natick, MA).

## Mutual Information

Mutual information (MI) quantifies information (in bit) shared between two variables, i.e.: how much the uncertainty about one variable (motion direction condition, in our case) may be reduced with knowledge about the second variable (spike counts). MI provides a robust metric of direction selectivity that accounts for changes in tuning curves, variability, and absolute firing rates. We calculated MI using:

$$MI(x,y) = \sum_x \sum_y p(x,y) \log_2 \left( \frac{p(x,y)}{p(x)p(y)} \right) \tag{1}$$

Where $p(x,y)$ is the joint probability distribution between number of spikes and direction condition, and $p(x)p(y)$ is the product of marginal probability distributions. This provides how many bits of information a neuron's firing rate contains about direction and vise versa. We calculated the mutual information between spike count and S1 direction for each non-overlapping 50ms bin from the start of each trial up to the end of the delay, and then the MI between spike count and S2 direction through the end of the trial. In order to correct for spurious effects that come from differences in trial-counts, we baseline-corrected MI by calculating it many times with shuffled conditions, and subtracted this out (Hatsopoulos et al. 1998). We then computed the bitrate by dividing this metric with the bin size (0.05s).

$$I(x;y) = \frac{MI(X;Y) - MI_{\text{shuffled}}(X;Y)}{0.05s} \tag{2}$$

## Classification of neurons as highly or weakly informative

We calculated baseline-corrected mutual information (MI) between the total number of spikes fired during the stimulus window and the direction condition. Neurons that were more informative about motion direction than the median (calculated across active and passive conditions combined) were classified as highly-informative, while those that fell below the median were weakly-informative.

## Entropy Factor

Entropy Factor (HF) quantifies the entropy of a neuron's spike count (across trials) relative to a Poisson process with the same rate parameter, $\lambda$ (Rajdl, Lansky, and Kostal 2017). It is similar to Fano factor, in that a value of 1 indicates equivalence to a Poisson process, while values greater or less than 1 indicate distributions that are supra-poisson or sub-poisson, respectively. However, unlike Fano factor, which compares the variance of a distribution of spike counts to its mean, entropy factor takes into account the entire distribution of spike counts. This provides a more reliable metric than Fano Factor, particularly when there are a low number of observations, for which sample variance is biased lower than the expected variance in the limit of infinite data. For each 50ms time bin, $t$, we calculated the entropy $H_{Observed}(\vec{x}_t)$ of the distribution of spike counts $(\vec{x}_t)$ across trials with the same stimulus direction and divided it by the entropy of a simulated Poisson distribution with the same intensity and number of trials.

$$HF(\vec{x}_t) = \frac{H_{Observed}(\vec{x}_t)}{H_{Poisson}(\hat{x})} \quad (3)$$

Here, $H_{Observed}(\vec{x}_t)$ is the entropy of the distribution of spike counts across trials $i \in [1, n]$ for a neuron at time $t$, defined by:

$$H_{Observed}(\vec{x}_t) = -\sum_{i=1}^{n} p(x_t^i) \log\left(p(x_t^i)\right) \quad (4)$$

And $H_{Poisson}(\hat{x})$ is the entropy of a simulated Poisson process with rate parameter $\lambda$:

$$\lambda = \bar{x}_t \quad (5)$$

The entropy of a Poisson process was calculated by doing 200 simulations of spike counts for $n$ trials pulled from a Poisson distribution with rate parameter $\lambda_t$, and the same unbiased bins used for the data. By then averaging the observed entropies of the simulations, we get an estimate of how variable, on average, we would expect the matched Poisson process to be. This normalization accounts for bias in sample entropy due to differences in firing rate or trial count.

## Tuning Curves

Tuning curves were calculated as the average firing rate during the stimulus window for each of the 8 directions. We then normalized the firing rates by the greatest response to any direction (preferred direction). For testing the effect of stimulus (S1 vs. S2) on tuning and HF, we used a three-way ANOVA with factors of 'direction' (eight levels corresponding to each motion direction), 'stimulus' (two levels: S1 vs. S2) and 'task' (two levels: active vs. passive), where each observation was the normalized firing rate for a neuron. We considered the effect of stimulus to be negligible if the main effect of 'stimulus' was not significant, and then combined data from S1 and S2 conditions for visualization purposes.

## Modulation Index

In order to calculate the modulation index, we first needed to ensure that the same neuron was recorded across tasks. The criteria used to determine if a neuron was matched were that (1) they were recorded on the same channel, (2) they had the same preferred direction of motion, and (3) their spike waveforms were highly correlated ($> 0.98$). This resulted in a set of 117 neurons that were recorded in both tasks. Modulation index was calculated as a neuron's response to a specific direction condition in active, minus the response in passive, divided by their sum.

$$ModulationIndex = \frac{ActiveFR - PassiveFR}{ActiveFR + PassiveFR} \quad (6)$$

This provides an estimate of response gain for each direction condition and each neuron that was recorded in both tasks.

## Comparison Effects

Comparison effects (CE) were calculated as the area under the curve of the Receiver Operating characteristic (AUC). For each neuron, trials were separated by direction condition, and then by trial type (same direction or different direction trials depending on whether S1 matched S2). CE was calculated on firing rates for each non-overlapping 50ms bin, as the AUC between 'same' trials and 'different' trials within direction. AUC values were then averaged across direction condition within each

time bin, resulting in a time course of AUC values for each neuron. Values greater than 0.5 indicates a neuron that responds more robustly to identical visual stimuli on trials where the directions of S1 and S2 matched ($S > D$ neurons), while values less than 0.5 indicated the opposite ($D > S$ neurons).

*Statistical Significance of CE* Statistical significance of CE was determined using a bootstrapping procedure. Trial labels ('same' or 'different' trial) were randomly permuted, and AUC values were calculated 200 times per neuron. If the average AUC during the post-S2 window was significantly different $\alpha = 0.01$ than this bootstrap distribution, it was considered significantly CE signalling. Above the distribution is same-preferring ($S > D$) and below is different-preferring ($D > S$).

## Noise Correlations

We calculated spike count correlations for pairs of neurons within each direction condition to measure stimulus-independent covariability ("noise" correlations). Spike count noise correlations were calculated separately for the 500ms stimulus windows. Pairwise Pearson product-moment correlations were calculated between simultaneously recorded neurons, transformed using the Fisher's R-to-Z transform, and then averaged across condition for each window, and transformed back into correlations. This provided an average noise correlation value between each pair of simultaneously-recorded neurons. These average pairwise noise correlations were then binned across recording sessions according the angular difference of the pairs' preferred direction of motion. Pairs were categorized as informative-pairs and mixed pairs depending on if one or both neurons in the pair classified as highly-informative about motion direction.

## Divisive Normalization Model

We adapted a divisive normalization model (Reynolds and Heeger 2009) in order to explain differences in attention modulation between highly and weakly informative neurons, as well as S1 and S2. The model was adapted to incorporate a heterogeneous neural population with different tuning to a second feature dimension. The model consists of 18 free parameters to describe two neural populations. These included: tuning width to two features (direction and a second, hypothetical, feature) for both populations (4 parameters), suppressive field widths for the two features and two neural populations (4 parameters), attention field widths for the two populations during S1 and S2 (4 parameters), baseline terms for the two populations during S1 and S2 (5 parameters), and the preferred value of the weakly-informative neurons along the invariant feature dimension (1 parameter). Parameters in the model were held constant across S1 and S2 with the exception of the attention field ($\theta$ width and $x$ width) and the baseline terms. We optimized the parameters using the matlab built-in fminsearch, which uses simplex search, with a standard root mean squared error (RMSE) function.

| | Parameter | Value | Fit to Data | Notes |
|---|---|---|---|---|
| Excitation | θ - HI | 0° | | HI = Highly-informative |
| | θ width - Hi | 57° | x | WI = Weakly-informative |
| | x - HI | 0 A.U. | | |
| | x width - HI | 15 A.U. | x | |
| | θ - WI | 0° | | |
| | θ width- WI | 27° | x | |
| | x - WI | 46 A.U. | x | |
| | x width - WI | 15 A.U. | x | |
| Supression | θ - HI | 0° | | |
| | θ width - HI | 1500° | x | |
| | x - HI | 0 A.U. | | |
| | | 50 A.U. | x | |
| | θ - WI | 0° | | |
| | θ width- WI | 120° | x | |
| | x - WI | 46 A.U. | | centered at WI excitation field x |
| | x width - WI | 120 A.U. | x | |
| Attention | θ S1 | 0° | | |
| | θ width S1 | 150° | x | |
| | θ S2 | 0° | | |
| | θ width S2 | 2000° | x | |
| | x S1 | 0 A.U. | | |
| | x width S1 | 46 A.U. | x | |
| | x S2 | 0 A.U. | | |
| | x width S2 | 58 A.U. | x | |
| Baseline | S1 HI | -0.00042 A.U. | x | |
| | S2 HI | 0.003 A.U. | x | |
| | S1 WI | -0.00045 A.U. | x | |
| | S2 WI | -0.00055 A.U. | x | |
| | post-normalization | 25 A.U. | x | |

**Figure 8.** *Model parameter values.* $\theta$ represents motion direction parameters. $x$ is the hypothetical second feature parameters.

## References


Arandia-Romero, Iñigo et al. (2016). "Multiplicative and Additive Modulation of Neuronal Tuning with Population Activity Affects Encoded Information". In: *Neuron* 89.6, pp. 1305–1316. ISSN: 10974199. DOI: 10.1016/j.neuron.2016.01.044.

Averbeck, Bruno B., Peter E. Latham, and Alexandre Pouget (2006). "Neural correlations, population coding and computation". In: *Nature Reviews Neuroscience* 7.5, pp. 358–366. ISSN: 1471003X. DOI: 10.1038/nrn1888.

Barbas, H. (1988). "Anatomic organization of basoventral and mediodorsal visual recipient prefrontal regions in the rhesus monkey". In: *Journal of Comparative Neurology* 276.3, pp. 313–342. ISSN: 10969861. DOI: 10.1002/cne.902760302.

Bartolo, Ramon et al. (2020). "Information-limiting correlations in large neural populations". In: *Journal of Neuroscience* 40.8, pp. 1668–1678. ISSN: 15292401. DOI: 10.1523/JNEUROSCI.2072-19.2019.

Bisley, James W. and Tatiana Pasternak (2000). "The multiple roles of visual cortical areas MT/MST in remembering the direction of visual motion". In: *Cerebral Cortex* 10.11, pp. 1053–1065. ISSN: 10473211. DOI: 10.1093/cercor/10.11.1053.

Bisley, James W., Daniel Zaksas, and Tatiana Pasternak (2001). "Microstimulation of cortical area MT affects performance on a visual working memory task". In: *Journal of Neurophysiology* 85.1, pp. 187–196. ISSN: 00223077. DOI: 10.1152/jn.2001.85.1.187.

Bisley, James W. et al. (2004). "Activity of Neurons in Cortical Area MT during A Memory for Motion Task". In: *Journal of Neurophysiology* 91.1, pp. 286–300. ISSN: 00223077. DOI: 10.1152/jn.00870.2003.

Born, Richard T. and David C. Bradley (2005). "Structure and function of visual area MT". In: *Annual Review of Neuroscience* 28, pp. 157–189. ISSN: 0147006X. DOI: 10.1146/annurev.neuro.26.041002.131052.

Britten, K. H. et al. (1992). "The analysis of visual motion: A comparison of neuronal and psychophysical performance". In: *Journal of Neuroscience* 12.12, pp. 4745–4765. ISSN: 02706474. DOI: 10.1523/jneurosci.12-12-04745.1992.

Carandini, Matteo and David J. Heeger (2012). "Normalization as a canonical neural computation". In: *Nature Reviews Neuroscience* 13.1, pp. 51–62. ISSN: 1471003X. DOI: 10.1038/nrn3136.

Churchland, Mark M. et al. (2010). "Stimulus onset quenches neural variability: A widespread cortical phenomenon". In: *Nature Neuroscience* 13.3, pp. 369–378. ISSN: 10976256. DOI: 10.1038/nn.2501. URL: http://dx.doi.org/10.1038/nn.2501.

Cohen, Marlene R. and John H.R. Maunsell (2009). "Attention improves performance primarily by reducing interneuronal correlations". In: *Nature Neuroscience* 12.12, pp. 1594–1600. ISSN: 10976256. DOI: 10.1038/nn.2439. arXiv: NIHMS150003.

Cohen, Marlene R. and William T. Newsome (2008). "Context-Dependent Changes in Functional Circuitry in Visual Area MT". In: *Neuron* 60.1, pp. 162–173. ISSN: 08966273. DOI: 10.1016/j.neuron.2008.08.007.

DeAngelis, Gregory C. and Takanori Uka (2003). "Coding of horizontal disparity and velocity by MT neurons in the alert macaque". In: *Journal of Neurophysiology*

89.2, pp. 1094–1111. ISSN: 00223077. DOI: 10.1152/jn.00717.2002.

Dias, Elisa C. et al. (2006). "Changing plans: neural correlates of executive control in monkey and human frontal cortex". In: *Experimental Brain Research* 174.2, pp. 279–291. ISSN: 00144819. DOI: 10.1007/s00221-006-0444-4.

Ding, Long (2015). "Distinct dynamics of ramping activity in the frontal cortex and caudate nucleus in monkeys". In: *Journal of Neurophysiology* 114.3, pp. 1850–1861. ISSN: 15221598. DOI: 10.1152/jn.00395.2015.

Durston, S. et al. (2002). "The effect of preceding context on inhibition: An event-related fMRI study". In: *NeuroImage* 16.2, pp. 449–453. ISSN: 10538119. DOI: 10.1006/nimg.2002.1074.

Gilbert, Charles D. and Wu Li (2013). "Top-down influences on visual processing". In: *Nature Reviews Neuroscience* 14.5, pp. 350–363. ISSN: 1471003X. DOI: 10.1038/nrn3476.

Hatsopoulos, N. G. et al. (1998). "Information about movement direction obtained from synchronous activity of motor cortical neurons". In: *Proceedings of the National Academy of Sciences of the United States of America* 95.26, pp. 15706–15711. ISSN: 00278424. DOI: 10.1073/pnas.95.26.15706.

Hussar, Cory and Tatiana Pasternak (2010). "Trial-to-trial variability of the prefrontal neurons reveals the nature of their engagement in a motion discrimination task". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.50, pp. 21842–21847. ISSN: 00278424. DOI: 10.1073/pnas.1009956107.

Hussar, Cory R. and Tatiana Pasternak (2009). "Flexibility of Sensory Representations in Prefrontal Cortex Depends on Cell Type". In: *Neuron* 64.5, pp. 730–743. ISSN: 08966273. DOI: 10.1016/j.neuron.2009.11.018. URL: http://dx.doi.org/10.1016/j.neuron.2009.11.018.

— (2012). "Memory-guided sensory comparisons in the prefrontal cortex: Contribution of putative pyramidal cells and interneurons". In: *Journal of Neuroscience* 32.8, pp. 2747–2761. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.5135-11.2012.

— (2013). "Common rules guide comparisons of speed and direction of motion in the dorsolateral prefrontal cortex". In: *Journal of Neuroscience* 33.3, pp. 972–986. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.4075-12.2013.

Katz, Leor N. et al. (2016). "Dissociated functional significance of decision-related activity in the primate dorsal stream". In: *Nature* 535.7611, pp. 285–288. ISSN: 14764687. DOI: 10.1038/nature18617.

Kaufman, Matthew T. et al. (2015). "Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex". In: *eLife* 4.MAY, pp. 1–21. ISSN: 2050084X. DOI: 10.7554/eLife.04677.

Khanna, Sanjeev B., Adam C. Snyder, and Matthew A. Smith (2019). "Distinct sources of variability affect eye movement preparation". In: *Journal of Neuroscience* 39.23, pp. 4511–4526. ISSN: 15292401. DOI: 10.1523/JNEUROSCI.2329-18.2019.

Kim, Hyung Goo R., Dora E. Angelaki, and Gregory C. DeAngelis (2015). "A functional link between MT neurons and depth perception based on motion parallax". In: *Journal of Neuroscience* 35.6, pp. 2766–2777. ISSN: 15292401. DOI: 10.1523/JNEUROSCI.3134-14.2015.

Kohn, Adam and J. Anthony Movshon (2003). "Neuronal adaptation to visual motion in area MT of the macaque". In: *Neuron* 39.4, pp. 681–691. ISSN: 08966273. DOI: 10.1016/S0896-6273(03)00438-0.

Kohn, Adam and Matthew A. Smith (2005). "Stimulus dependence of neuronal correlation in primary visual cortex of the macaque". In: *Journal of Neuroscience* 25.14, pp. 3661–3673. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.5106-04.2005.

Luck, Steven J. et al. (1997). "Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex". In: *Journal of Neurophysiology* 77.1, pp. 24–42. ISSN: 00223077. DOI: 10.1152/jn.1997.77.1.24. URL: http://www.physiology.org/doi/10.1152/jn.1997.77.1.24.

Lui, Leo L. and Tatiana Pasternak (2011). "Representation of comparison signals in cortical area MT during a delayed direction discrimination task". In: *Journal of Neurophysiology* 106.3, pp. 1260–1273. ISSN: 00223077. DOI: 10.1152/jn.00016.2011.

Martinez-Trujillo, Julio and Stefan Treue (2004). "Feature-Based Attention Increases the Selectivity of Population Responses in Primate Visual Cortex". In: *Current Biology* 14, p. 1118. DOI: 10.1016/j.

Maunsell, J. H.R. and D. C. Van Essen (1983). "Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation". In: *Journal of Neurophysiology* 49.5, pp. 1148–1167. ISSN: 00223077. DOI: 10.1152/jn.1983.49.5.1148.

Mitchell, Jude F., Kristy A. Sundberg, and John H. Reynolds (2007). "Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4". In: *Neuron* 55.1, pp. 131–141. ISSN: 08966273. DOI: 10.1016/j.neuron.2007.06.018.

Moreno-Bote, Rubén et al. (2014). "Information-limiting correlations". In: *Nature Neuroscience* 17.10, pp. 1410–1417. ISSN: 15461726. DOI: 10.1038/nn.3807.

Narayanan, Nandakumar S. (2016). "Ramping activity is a cortical mechanism of temporal control of action". In: *Current Opinion in Behavioral Sciences* 8, pp. 226–230. ISSN: 23521546. DOI: 10.1016/j.cobeha.2016.02.017.

Panzeri, Stefano et al. (1999). "Correlations and the encoding of information in the nervous system". In: *Proceedings of the Royal Society B: Biological Sciences* 266.1423, pp. 1001–1012. ISSN: 14712970. DOI: 10.1098/rspb.1999.0736.

Pasternak, Tatiana and Duje Tadin (2020). "Linking Neuronal Direction Selectivity to Perceptual Decisions about Visual Motion". In: *Annual Review of Vision Science* 6, pp. 335–362. ISSN: 23744650. DOI: 10.1146/annurev-vision-121219-081816.

Petrides, M. and D. N. Pandya (2006). "Efferent Association Pathways Originating in the Caudal Prefrontal Cortex in the Macaque Monkey". In: *Journal of Comparative Neurology* 498.October 2007, pp. 227–251. DOI: 10.1002/cne.

Ponce-Alvarez, Adrián et al. (2013). "Stimulus-dependent variability and noise correlations in cortical MT neurons". In: *Proceedings of the National Academy of Sciences of the United States of America* 110.32, pp. 13162–13167. ISSN: 00278424. DOI: 10.1073/pnas.1300098110.

Quian Quiroga, Rodrigo and Stefano Panzeri (2009). "Extracting information from neuronal populations: Information theory and decoding approaches". In: *Nature Reviews Neuroscience* 10.3, pp. 173–185. ISSN: 1471003X. DOI: 10.1038/nrn2578.

Rajdl, K., P. Lansky, and L. Kostal (2017). "Entropy factor for randomness quantification in neuronal data". In: *Neural Networks* 95, pp. 57–65. ISSN: 18792782. DOI: 10.1016/j.neunet.2017.07.016.

Renart, Alfonso and Christian K. Machens (2014). "Variability in neural activity and behavior". In: *Current Opinion in Neurobiology* 25, pp. 211–220. ISSN: 18736882. DOI: 10.1016/j.conb.2014.02.013.

Reynolds, John H. and David J. Heeger (2009). "The Normalization Model of Attention". In: *Neuron* 61.2, pp. 168–185. ISSN: 08966273. DOI: 10.1016/j.neuron.2009.01.002. URL: http://dx.doi.org/10.1016/j.neuron.2009.01.002.

Rodman, Hillary R. and Thomas D. Albright (1987). "Coding of visual stimulus velocity in area MT of the macaque". In: *Vision Research* 27.12, pp. 2035–2048. ISSN: 00426989. DOI: 10.1016/0042-6989(87)90118-0.

Rudolph, Kirsten and Tatiana Pasternak (1999). "Transient and permanent deficits in motion perception after lesions of cortical areas MT and MST in the macaque monkey". In: *Cerebral Cortex* 9.1, pp. 90–100. ISSN: 10473211. DOI: 10.1093/cercor/9.1.90.

Ruff, Douglas A and Marlene R Cohen (2016). "Attention Increases Spike Count Correlations between Visual Cortical Areas". In: 36.28, pp. 7523–7534. DOI: 10.1523/JNEUROSCI.0610-16.2016.

Salzman, C. D. et al. (1992). "Microstimulation in visual area MT: Effects on direction discrimination performance". In: *Journal of Neuroscience* 12.6, pp. 2331–2355. ISSN: 02706474. DOI: 10.1523/jneurosci.12-06-02331.1992.

Salzman, C. Daniel, Kenneth H. Britten, and William T. Newsome (1990). "Cortical microstimulation influences perceptual judgements of motion direction". In: *Nature* 346.6280, pp. 174–177. ISSN: 00280836. DOI: 10.1038/346174a0.

Semedo, João D. et al. (2019). "Cortical Areas Interact through a Communication Subspace". In: *Neuron* 102.1, 249–259.e4. ISSN: 10974199. DOI: 10.1016/j.neuron.2019.01.026.

Shadlen, Michael N. and William T. Newsome (2001). "Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey". In: *Journal of Neurophysiology* 86.4, pp. 1916–1936. ISSN: 0022-3077. DOI: 10.1152/jn.2001.86.4.1916. URL: http://www.physiology.org/doi/10.1152/jn.2001.86.4.1916.

Smith, Matthew A. and Adam Kohn (2008). "Spatial and temporal scales of neuronal correlation in primary visual cortex". In: *Journal of Neuroscience* 28.48, pp. 12591–12603. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.2929-08.2008.

Smith, Matthew A. and Marc A. Sommer (2013). "Spatial and temporal scales of neuronal correlation in visual area V4". In: *Annals of Internal Medicine* 158.6, pp. 5422–5432. ISSN: 00034819. DOI: 10.1523/JNEUROSCI.4782-12.2013.

Snyder, Adam C., Byron M. Yu, and Matthew A. Smith (2018). "Distinct population codes for attention in the absence and presence of visual stimulation". In: *Nature Communications* 9.1. ISSN: 20411723. DOI: 10.1038/s41467-018-06754-5.

Snyder, Adam C., Byron M. Yu, and Matthew A. Smith (2021). "A Stable Population Code for Attention in Prefrontal Cortex Leads a Dynamic Attention Code in Visual Cortex". In: *The Journal of Neuroscience* 41.44, pp. 9163–9176. ISSN: 0270-6474. DOI: 10.1523/jneurosci.0608-21.2021.

Ungerleider, Leslie G. and Robert Desimone (1986). "Cortical connections of visual area MT in the macaque". In: *Journal of Comparative Neurology* 248.2, pp. 190–222. ISSN: 10969861. DOI: 10.1002/cne.902480204.

Wimmer, K. et al. (2016). "Transitions between Multiband Oscillatory Patterns Characterize Memory-Guided Perceptual Decisions in Prefrontal Circuits". In: *Journal of Neuroscience* 36.2, pp. 489–505. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.3678-15.2016. URL: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3678-15.2016.

Zaksas, Daniel and Tatiana Pasternak (2006). "Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task". In: *Journal of Neuroscience* 26.45, pp. 11726–11742. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.3420-06.2006.