RESEARCH ARTICLE

Functional specialization in the middle temporal area for smooth pursuit initiation [version 2; peer review: 2 approved]

Shahab Bakhtiari, Christopher C. Pack
Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

Abstract
Smooth pursuit eye movements have frequently been used to model sensorimotor transformations in the brain. In particular, the initiation phase of pursuit can be understood as a transformation of a sensory estimate of target velocity into an eye rotation. Despite careful laboratory controls on the stimulus conditions, pursuit eye movements are frequently observed to exhibit considerable trial-to-trial variability. In theory, this variability can be caused by the variability in sensory representation of target motion, or by the variability in the transformation of sensory information to motor commands. Previous work has shown that neural variability in the middle temporal (MT) area is likely propagated to the oculomotor command, and there is evidence to suggest that the magnitude of this variability is sufficient to account for the variability of pursuit initiation. This line of reasoning presumes that the MT population is homogeneous with respect to its contribution to pursuit initiation. At the same time, there is evidence that pursuit initiation is strongly linked to a subpopulation of MT neurons (those with strong surround suppression) that collectively generate less motor variability. To distinguish between these possibilities, we have combined human psychophysics, monkey electrophysiology, and computational modeling to examine how the pursuit system reads out the MT population during pursuit initiation. We find that the psychophysical data are best accounted for by a model that gives stronger weight to surround-suppressed MT neurons, suggesting that variability in the initiation of pursuit could arise from multiple sources along the sensorimotor transformation.

Keywords
Smooth pursuit, area MT, variability, noise correlation, surround suppression, computational modeling

Open Peer Review

Reviewer Status
Invited Reviewers
1
2

Accepted

version 2
published
18 Mar 2019

version 1
published
31 Oct 2018

1 Gunnar Blohm, Queen's University, Kingston, Canada

2 Nicholas S. C. Price, Monash University, Clayton, Australia

Joanita D'Souza, Monash University, Clayton, Australia

Any reports and responses or comments on the article can be found at the end of the article.
Introduction

When an object of interest moves through the visual field, it can be tracked by an eye rotation that matches the target velocity. Such smooth pursuit eye movements serve to stabilize the image of a moving object on the retina (Robinson et al., 1986). In the brain, the dynamics of smooth pursuit are controlled by a complex interplay of sensory and motor signals (Krauzlis & Lisberger, 1994; Lisberger, 2015; Thier & Ilg, 2005). However, the initiation phase of pursuit is relatively simple to understand, as the eye velocity during this time is mainly determined by the neural representation of the target motion. Therefore, smooth pursuit initiation can be modeled as a transformation of sensory information about target motion to a motor command that rotates the eye at the appropriate velocity (Ilg, 2008; Lisberger, 2010). Given its relatively well-known underlying neuronal basis, and the availability of precise eye tracking, smooth pursuit initiation provides a powerful model of neural coding.

Models of neural coding generally consider two key factors: the stimulus sensitivity of individual neurons and the noise correlations among neurons (Zohary et al., 1994). Low sensitivity or high levels of correlated noise can lead to large variability in stimulus coding (see below), which in turn limits the accuracy and precision of behavioral performance. For smooth pursuit, previous work has shown that the gain of pursuit initiation can be predicted from the properties of neurons in the middle temporal (MT) area of macaque visual cortex (Priebe & Lisberger, 2004). MT neurons exhibit high sensitivity to stimulus velocity, and their average firing rates for stimuli of different spatial frequency or contrast are predictive of smooth pursuit velocity for the same stimuli. Similarly, computational models have been able to link the trial-to-trial variability of smooth pursuit initiation to the correlated variability of MT neuronal activity (Hohl et al., 2013; Huang & Lisberger, 2009; Lee et al., 2016; Osborne et al., 2005). These observations suggest that variability in smooth pursuit initiation is largely due to the population activity in MT.

Quantitative estimates of the relationship between behavioral variability and neural variability have typically assumed that the MT population is homogeneous with respect to its contribution to smooth pursuit initiation (Huang & Lisberger, 2009). However, it has also been argued that pursuit initiation is primarily tied to a subset of neurons that can be detected based on their center-surround receptive field organization. In particular, the results of microstimulation experiments suggest that MT neurons with strong suppressive surrounds contribute causally to estimates of target velocity, while neurons with weaker surround suppression encode the retinal motion of the background (Born et al., 2000). Moreover, the structure of noise correlation is different in the population of these two types of MT neurons. Namely, the neurons with strong surround suppression have a noise correlation structure that is less detrimental to population coding of the stimulus (Liu et al., 2016). Indeed, the difference between the two classes of neurons in noise correlation can have crucial implications for the relation between pursuit and MT variability. Because these classes of neurons form different projections to downstream cortical areas (Berezovskii & Born, 2000), it could be useful to consider their properties separately in formulating hypotheses about the relationship between neural activity in MT and pursuit behavior.

In terms of sensitivity, surround-suppressed neurons are by definition less responsive to stimuli that are larger or smaller than their classical receptive field sizes (Allman et al., 1985). In contrast, neurons that lack surround suppression show responses that are invariant over a broad range of stimulus sizes. These considerations lead to specific predictions about the properties of smooth pursuit initiation. If pursuit is driven primarily by surround-suppressed MT neurons, its variability should increase for large stimuli, as the ability of this subpopulation to signal motion decreases (Liu et al., 2016). On the other hand, if pursuit is driven by the MT population as a whole, its variability should decrease or plateau with for large stimuli, as they activate more MT neurons (Shadlen et al., 1996). In this study, we have examined the validity of these hypotheses through psychophysical experiments and computational modeling. In order to probe the importance of surround suppression for pursuit initiation and its variability, we conducted experiments in which the size of the smooth pursuit target changed from trial to trial.

Our psychophysical and computational results support the idea that pursuit initiation accords greater weight to MT neurons with stronger surround suppression (Born et al., 2000). Because this population of neurons exhibits correlated variability that is unlikely to strongly affect stimulus coding (Liu et al., 2016), these results suggest that additional variability in pursuit velocity is introduced downstream of MT.

Methods

Experimental procedure

We used a paradigm similar to that of Hohl et al. (2013) to test the influence of stimulus size on smooth pursuit initiation. Stimuli were generated in MATLAB with the Psychophysics Toolbox (Pelli, 1997) and presented on an OLED TV (LG 55EA9800, 122.7 cm × 79.86 cm) at a frame rate of 120 Hz. Each pixel subtended 0.067° of visual angle at a viewing distance of 57 cm, and the screen covered 47° and 35° of horizontal and vertical visual angle. The experiments were carried out in total darkness.
A total of seven subjects, five females and two males, mean age 26.8 years (range 20–34), with normal vision were recruited for this experiment. Subjects, except one of the authors (SB), were recruited via online announcements posted on Student’s Society of McGill University (SSMU) Market Place. Except for one of the subjects (author SB), subjects were naive and unexperienced in psychophysics experiments. Chin and forehead rests were used to stabilize the subjects’ heads and to maximize the accuracy of eye tracking. All subjects gave written informed consent to participate in the study, which was approved by the Ethics Committee of the Montreal Neurological Institute and Hospital (NEU-06-033).

Figure 1A contains a schematic representation of the task paradigm. Each trial started with a red fixation point subtending 0.2° (luminance 55.2 cd/m²), positioned in the center of the screen. Simultaneous with the fixation point, a red circle subtending 1°, with the same luminance as the fixation point, appeared at 10° in the periphery. The peripheral circle was the main target of smooth pursuit. After a fixation period chosen randomly between 500 and 1500 ms, the main target became dim (luminance 1.5 cd/m²) and a patch of small white dots (each dot 0.1° of visual angle, 0.3% density, 100% coherence, luminance 75 cd/m²) appeared around the main target (Hohl et al., 2013). The low luminance of the target during this time period was chosen to maximize the influence.
of random dots on smooth pursuit initiation. The patch of random dots and the main target had the same direction, chosen randomly to be either leftward or rightward, and the same speed, which was chosen on each trial randomly between 5° and 20° s⁻¹.

150 ms after the onset of target motion, the patch of random dots disappeared, while the main target continued with the same velocity for another 750 milliseconds. The subject had to saccade toward the moving target (red, low contrast circle) as soon as he/she detected the motion, and follow it with post-saccadic smooth pursuit. The task had three conditions, with the only difference between the conditions being the size of the patch of random dots (2, 6, or 20 degrees). All the stimuli were presented on a dark background.

Eye movement data analysis
Eye position was recorded using a desk-mounted EyeLink 1000 eye tracker (SR Research) at 1000 Hz. Recordings were monocular (right eye only), and the data were analyzed offline. Saccades were detected from the recorded eye positions using an automated acceleration-based algorithm (Groh et al., 1997), followed by visual inspection to correct false detections. On each trial, the period of 20–80 ms after the saccade was chosen as the open-loop phase of smooth pursuit (Born et al., 2000). After low-pass filtering (0.1–25 Hz), the average velocity of the eye during this period was calculated and passed to subsequent stages for further analysis.

Statistical analysis
The relationship between eye velocity and target velocity was studied with a linear model. Separate models were fitted to the data from each individual subject. The models had the form:

\[ y_j = \alpha_j + \beta_j x_j + \epsilon_j, \]  

(1)

where \( y \) and \( x \) represent the eye and target velocities, respectively; \( \alpha \) is the intercept, \( \beta \) is the slope of the fitted linear model, and \( \epsilon \) is the error term with distribution \( \mathcal{N}(0,\sigma) \) for the \( j \)th subject. In this formulation, the slope and the intercept can vary across subjects. The average of the normalized estimated parameters across the subjects is the intercept, \( \hat{\alpha} \), and the slope, \( \hat{\beta} \), were used for group-level analysis.

We also fitted a linear mixed effect model to the data from all subjects pooled together. The model had the form:

\[ y = \alpha_j + \beta_j x + \epsilon, \]  

where \( y \) and \( x \) represent the eye and target velocities from all the subjects, respectively, with the target velocity being the fixed effect. As with the individual-subject models, this model can have varying slope and intercept; the subject identity is the random effect. \( \epsilon \) is the error term with distribution \( \mathcal{N}(0,\sigma) \).

The separately fitted linear models and the mixed effect model produced similar results. Therefore, here we only report the results of the separately fitted linear models, unless stated otherwise.

We used the Statistics and Machine Learning Toolbox™ in MATLAB 2016 for fitting the linear and the mixed-effect models.

Computational model
To interpret our behavioral findings, we extended an existing model of MT (Hohl et al., 2013) to include recent observations on the relationship between surround suppression and noise correlations (Liu et al., 2016). Except where noted below, all parameters were constrained either by the existing model (Hohl et al., 2013) or by MT recordings (Liu et al., 2016).

Simulation of MT population. We simulated a population of MT neurons with parameters related to direction, speed, and size tuning (Hohl et al., 2013). For each cell, the preferred velocity (\( \theta_{\text{pref}} \)) and direction (\( \theta_{\text{null}} \)) were sampled randomly from 0.5–512 degrees/second and 0–180 degrees, respectively. Our method for characterizing the size tuning curves for MT neurons is detailed in (Liu et al., 2016). Briefly, for each stimulus size, we first calculate the neuronal \( d' \) as:

\[ d' = \frac{r_{\text{pref}} - r_{\text{null}}}{\sqrt{\sigma_{\text{pref}}^2 + \sigma_{\text{null}}^2}} \]  

(2)

where \( r_{\text{pref}} \) and \( r_{\text{null}} \) are the mean responses to the preferred and null directions, and \( \sigma_{\text{pref}}^2 \) and \( \sigma_{\text{null}}^2 \) are the response variances. The resulting size tuning curves are then fitted with the Difference of Error Functions (DoE) function (Equation (4) below) (Liu et al., 2016).

The following equations determine the mean firing rate of each cell:

\[ f_i(\theta) = e^{-0.5 \left( \frac{\log(\theta_{\text{null}}) - \log(\theta_{\text{pref}})}{\sigma} \right)^2} \]  

(3)

\[ f_i(\rho) = A_e e^{\rho} - A_k e^{\frac{\rho}{\sigma}} \]  

(4)

\[ R(\theta, \theta_{\text{null}}) = R_0 + G f_i(\theta).f_i(\theta_{\text{null}}) \]  

(5)

where the speed, direction, and size tuning are respectively given by \( f_i(\theta) \), \( f_i(\theta_{\text{null}}) \), and \( f_i(\rho) \). \( R_0 \) indicates the baseline mean firing rate, and \( G \) represents the neuron’s gain. \( (\theta, \theta_{\text{null}}, \rho) \) are the stimulus parameters (speed, direction, and size.) \( \sigma \) and \( \sigma_{\text{null}} \) also determine the tuning curve widths, and \( \sigma \) and \( \sigma_{\text{null}} \) are excitatory and inhibitory sizes. \( A_e \) and \( A_k \) also determine the excitatory and inhibitory heights, respectively. Except for the size tuning parameters, which have been sampled from our MT electrophysiology recordings, all the other parameters are chosen based on the values used in (Hohl et al., 2013). On each trial, the firing rates of the simulated MT neurons are sampled from a multivariate Gaussian distribution with the mean vector of \( R(\theta, \theta_{\text{null}}, \rho) \) and a covariance matrix calculated based on the method in (Shadlen et al., 1996). The correlation structure between the neurons is:

\[ \Delta_{\text{pref}} = \log(v_i^{\text{pref}}) - \log(v_i^{\text{null}}) \]  

(7)
\[ \Delta \theta^i_{\text{pref}} = \theta^i_{\text{pref}} - \theta^i \]  
(8)

where \( \Delta \theta^i_{\text{pref}} \) and \( \Delta \theta^i \) are calculated based on (7) and (8), but the preferred directions and speeds are replaced with random samples from the range of possible values.

An important extension to the model of Hohl and colleagues is the addition of a term \( \alpha^i \), which models the relationship between surround suppression and noise correlations observed by (Liu et al., 2016); its value is determined by the surround suppression indices of the \( i^\text{th} \) and \( j^\text{th} \) neurons. The surround suppression index for the \( i^\text{th} \) neuron is calculated as \( \text{SI}_i = (d_m^i - d_m^i)/d_m^i \), where \( d_m^i \) is the maximum response across the responses to different stimulus sizes, and \( d_m^i \) is the response to the largest stimulus size (Liu et al., 2016). For two neurons with strong surround suppression, \( \alpha^i \) will be close to one, and \( r^i_{\text{noise}} \) in (9) will not depend on the tuning similarity of the two neurons. The opposite is true for two cells with no surround suppression. This captures the observation (Liu et al., 2016) that pairs of surround-suppressed neurons have noise correlations that are largely independent of the signal correlations, which has important implications for neural coding (Averbeck et al., 2006; see Discussion below).

The surround suppression dependency \( (\alpha^i) \) can be defined in different forms, and indeed its definition affects the structure of noise correlation. We incorporated noise correlations measured from pairs of real MT neurons to obtain a more realistic modeling of surround suppression dependency in our model. We assumed four possible structures for \( \alpha^i \): 1) \( \alpha^i \) is directly proportional to the sum of the surround suppression indices of the two neurons,

\[ \alpha^i = \frac{\text{SI}_i + \text{SI}_j}{\max_j(\text{SI}_i, \text{SI}_j)} \]  
(10)

2) \( \alpha^i \) is equal to 0, 0.5, or 1, based on the surround suppression indices of the two neurons,

\[ \alpha^i = \begin{cases} 0, & \text{SI}_i \text{ and } \text{SI}_j < \text{median(SI)} \\ 0.5, & \text{SI}_i \geq \text{median(SI)} \text{ and } \text{SI}_j < \text{median(SI)} \\ 1, & \text{SI}_i \text{ and } \text{SI}_j \geq \text{median(SI)} \end{cases} \]  
(11)

3) \( \alpha^i \) is drawn randomly from Gaussian distributions with mean values of 0, 0.5, and 1, and standard deviations of 1. The mean value of the Gaussian distribution is chosen based on the surround suppression indices of the two neurons,

\[ \alpha^i \sim \begin{cases} \mathcal{N}(0, 1), & \text{SI}_i \text{ and } \text{SI}_j < \text{median(SI)} \\ \mathcal{N}(0.5, 1), & \text{SI}_i \geq \text{median(SI)} \text{ and } \text{SI}_j < \text{median(SI)} \\ \mathcal{N}(1, 1), & \text{SI}_i \text{ and } \text{SI}_j \geq \text{median(SI)} \end{cases} \]  
(12)

and 4) \( \alpha^i \) is equal to zero for all pairs of neurons which represents the structure in which noise correlation does not depend on surround suppression at all.

To choose between these four models, we compared their marginal likelihoods \( p(\tilde{\tau}^i \mid \mathcal{M}) \). Since our noise correlation models are parameterized by \( \tau_d \) and \( \tau_r \), the Bayesian model evidence is estimated by integrating over \( \tau_d \) and \( \tau_r \):

\[ p(\tilde{\tau}^i \mid \mathcal{M}) = \int p(\tilde{\tau}^i \mid \mathcal{M}) p(\mathcal{M} \mid \mathcal{M}) d\mathcal{M} \]  
(13)

where \( \mathcal{M} = (\tau_d, \tau_r) \) represents the \( k^\text{th} \) suggested model of surround suppression dependency as explained above, and \( k = 1, 2, 3, 4 \). \( \tilde{\tau}^i \) is sampled from (0,100) with a resolution of 4.5 for \( \tau_d \) and \( \tau_r \). In this comparison, the third model, described by (12), produced the largest marginal likelihood (see Results and Figure 6). The values of \( \tau_d \) and \( \tau_r \) were then chosen to maximize the likelihood

\[ p(\tilde{\tau}^i \mid \mathcal{M}) = \prod_{i=1}^N p(\tilde{\tau}^i \mid \mathcal{M}) \]  
(14)

where \( N \) is the number of noise correlation samples measured from area MT. The likelihood \( p(\tilde{\tau}^i \mid \mathcal{M}) \) was estimated via repeated simulation of the noise correlation under different values of \( \tau_d \) and \( \tau_r \). In our analysis, \( \tau_d = 45 \), and \( \tau_r = 4.5 \) maximized the likelihood function, and were used in our simulations.

Given the noise correlation \( \rho^i_{\text{noise}} \), the covariance matrix of the population is then defined based on the method suggested in Shadlen et al. (1996).

In our model, the number of MT neurons in the population that respond to a specific stimulus is determined by the size of the stimulus. The cortical magnification in area MT, Magnification Factor = 6 x eccentricity^0.9 (Erickson et al., 1989; Van Essen et al., 1981), is used to map visual space in degrees to cortical space in millimeters. Integrating over the cortical space based on this equation yields the cortical area (in square millimeters) that is being activated by a stimulus with a specific diameter. The absolute number of activated neurons can then be obtained by multiplying the activated cortical area by the number of neurons per square millimeter in area MT. In our simulations, we used 20 neurons/mm^2 to obtain population sizes that are consistent with previous studies (Liu et al., 2016).

Decoding stimulus velocity. Given the simulated MT population responses, vector averaging (Hohl et al., 2013; Lisberger & Ferrera, 1997) was used to decode the stimulus speed from the population activity:

\[ s_h^i = \frac{\sum_{p=1}^P \sum_{j=1}^J \sum_{i=1}^I W(p, i) W(j, i') \cos(\theta^i_{\text{pref}}) R(i') \log(\theta^i_{\text{pref}})}{\sum_{j=1}^J \sum_{i=1}^I W(p, i) W(j, i') R(i')} \]  
(15)

\[ s_v^i = \frac{\sum_{p=1}^P \sum_{j=1}^J \sum_{i=1}^I W(p, i) W(j, i') \sin(\theta^i_{\text{pref}}) R(i') \log(\theta^i_{\text{pref}})}{\sum_{j=1}^J \sum_{i=1}^I W(p, i) W(j, i') R(i')} \]  
(16)
We selected the parameters of the weight functions (i.e., $D$ in $W_p(.)$ and $a$ in $W_s(.)$) by comparing our behavioral data with the simulated pursuit velocity for a predetermined range of parameters. Specifically, as a measure of behavioral suppression, we used the observed ratio of pursuit variability at 20 degrees to the variability at 6 degrees (see Experimental Procedure).

$$\text{SI}_{\text{behav}} = \frac{\sigma_p^{20}}{\sigma_p^{6}}$$

where, $\sigma_p^{6}$ and $\sigma_p^{20}$ are smooth pursuit variabilities at 6 and 20 degrees of stimulus size, respectively for the $p$ subject. In the model, we simulated $\tilde{S}_i^{\text{behav}}$ on a grid of $a$ and $D$ values ($D \in \{4,8,16,32\}$, $a \in [0; 2; 20]$). By projecting the $\tilde{S}_i^{\text{behav}}$ values on the simulated $\tilde{S}_i^{\text{behav}}$ values, we chose the parameter values that generated the closest behavioral suppression to the average behavioral data. To this end, the 3 nearest neighbors for each $\sigma_p^{20}$ among the simulated $\tilde{S}_i^{\text{behav}}$ were detected, and the average of their corresponding $D$ and $a$ values were selected for the simulations.

**Analysis of neural data.** We re-analyzed previously published data recorded from monkey area MT (Liu et al., 2016). In these experiments, the motion stimulus was presented for 50 ms to the monkey while electrophysiology recording was being done simultaneously from area MT (for details see (Liu et al., 2016)). Using linear microelectrode arrays, in each session of experiments, four to seven single units were recorded simultaneously. We used the population of recorded neurons in each session to decode the direction of stimulus motion from the population activity in each trial. Specifically, we would like to see how the decoder’s certainty changes as a function of stimulus size.

To this end, we used two-class linear discriminant analysis (LDA) (Bishop, 2007) to decode the direction of motion from the activity of the neurons in each trial. Given the LDA prediction ($\hat{\theta}$) and the true stimulus motion direction ($\theta$) for each trial, we calculated the uncertainty coefficient of the decoder as:

$$\text{Uncertainty} = \frac{I(\theta, \hat{\theta})}{H(\theta)}$$

where, $I(\theta, \hat{\theta})$ is the mutual information of the decoded and real stimulus direction, and $H(\theta)$ is the entropy of the real stimulus direction (Cover & Thomas, 2012). The uncertainty coefficient quantifies how precisely the decoded direction predicts the real direction of the stimulus. Importantly, high uncertainty is assigned to a decoder which is inconsistent in decision making; wrong but consistent decisions do not result in high uncertainty. When sampling from a trial to another trial from the decisions of an inconsistent decoder (with high uncertainty coefficient) for a constant stimulus direction, one obtains more variable decisions, compared to those of a consistent decoder. In other words, the output of an uncertain decoder has high trial-to-trial variability. Therefore, as a proxy to behavioral variability, we measured the decoder uncertainty for different stimulus sizes. We then calculated the weighted average of uncertainties over the recording sessions. Once the weights assigned to each session was proportional to the number of surround suppressed neurons (SI $> \text{median(SI)}$) red curve in Figure 2A) and once proportional to the number of non-surround suppressed neurons (SI $< \text{median(SI)}$) (blue curve in Figure 2A). Comparison of these two weighted averages in Figure 2A shows the effect of surround suppression on the decoder uncertainty.

**Results**

We tested the effect of stimulus size on the precision of smooth pursuit initiation. As detailed in the Methods, changing the stimulus size influences motion encoding in MT population in two ways: first, by changing the pattern of activation across the MT population, and second, through the structure of the noise correlations (Liu et al., 2016). We hypothesized that the variability of smooth pursuit primarily reflects the properties of a subpopulation of MT neurons, namely those with strong surround suppression (Born et al., 2000). This hypothesis implies that increasing the stimulus size first increases the precision of the pursuit initiation, but then, due to the decreased sensitivity of the suppressed cells, it deteriorates the behavioral precision (see Figure 2 for a schematic presentation of our hypothesis). We verified this hypothesis with psychophysics experiments and computational modeling, as reported in the following sections.

**Analysis of neural data from MT**

To generate predictions based on the properties of different neuronal populations, we re-analyzed previously published data recorded from monkey area MT (Liu et al., 2016). Specifically, we decoded the direction of stimulus motion from groups of four to seven simultaneously recorded MT neurons. We then used an optimal decoder (see Methods) (Powers, 2011) to estimate the uncertainty associated with the neural representation of velocity in the MT population. As shown in Figure 2, the uncertainty varies with stimulus size, and this relationship is different for surround-suppressed and non-suppressed subpopulations. For a population of neurons with surround suppression, there is an

$$s = 2\sqrt{s_0^2 + s_c^2}$$

where

$$W_p(c_i) = \begin{cases} 1, & c_i < D \\ 0, & c_i \geq D \end{cases}$$

$$W_s(S_I) = \frac{1}{1 + e^{-a(S_I - \text{median(SI)})}}.$$
Figure 2. Direction decoding uncertainty from middle temporal (MT) population. (a) Direction decoding uncertainty. We decoded the direction of moving stimuli based on the responses of groups of simultaneously recorded MT neurons. The uncertainty of the decoder in assigning directions to the stimuli is plotted as a function of different stimulus sizes. The uncertainty of trained decoders are averaged once weighted based on the number of non-surround suppressed neurons in the sets (blue curve) and once based on the number of surround suppressed neurons (red curve). Therefore, the red curve shows the average decoder uncertainty when the majority of neurons in the population have surround suppression, while the blue curve shows the uncertainty for a population mainly including surround suppressed neurons. (b) A schematic representation of our hypothesis. The first column represents the stimulus in three different sizes. The second column schematically demonstrates the activity that each stimulus evokes in area MT. The small circles and the triangles show the neurons with strong and weak surround suppression, respectively. The dotted lines within the boxes separate the surround-suppressed and non-surround-suppressed neurons. The stimulus induced activity level is color coded and explained by the size tuning curves at top row. Blue and red colors represent low and high firing rates, respectively. The largest stimulus evokes the highest average level of activity in neurons without surround suppression (red triangles) while it induces a lower firing rate in surround suppressed cells (blue circles). The third column represents the decoder. $f(MT)$ estimates the velocity of the stimulus in the first column given the activity of MT population in each trial. According to one hypothesis, the neurons with strong surround suppression have a stronger contribution (solid gray arrow) in the readout compared to the neurons with weak surround suppression (dashed white arrow.) The trial-to-trial variability of the estimated velocity is depicted in the fourth row. The least variability, among these three stimulus sizes, is given by the medium sized stimulus (middle row) where the signal to noise ratio is higher than the two other conditions for the cells with surround suppression.
increase in the decoder’s uncertainty for large stimuli (Figure 2, red curve), while the uncertainty decreases and then plateaus with increasing stimulus size for the neurons with weaker surround suppression (Figure 2, blue curve).

Psychophysics experiments
In our psychophysics experiments, human subjects pursued a moving target that was surrounded by coherently moving dots (Figure 1A). For each subject and each trial, we quantified the first 60 ms of the post-saccadic smooth pursuit velocity. Figure 1B shows the raw eye position data for one subject, as well as the segment of the pursuit eye movement that was used for further pursuit velocity analysis (pink). Figure 1C shows examples of measured post-saccadic eye velocity.

In each trial of the task, a random velocity between 5 and 20 °/sec was assigned to the pursuit target. The scatter plot in Figure 3A shows the eye velocity as a function of target velocity on every trial in one subject.

In Figure 3A, the small gray circles represent the eye and target velocity in each trial. To measure the accuracy and precision of the

**Figure 3. Sample of regression analysis on one subject.** Top: The gray circles represent the single trials for one of the conditions (target size = 6 degrees.) The horizontal and the vertical axes show the target and the eye velocities, respectively. The green dashed line demonstrates the linear model fitted to the measured target/eye velocity pairs. Bottom: The Root Mean Square Error (RMSE) of the fitted lines in (a), for the three conditions. The error bars show the standard error on samples of RMSE values estimated by bootstrapping the data.
smooth pursuit system, we then fitted a linear model to the pairs, as explained in the Methods section. This process was repeated across the subjects and across task conditions (i.e. different sizes of random dots patches.)

For an ideal smooth pursuit system, we expect to obtain a linear fit with a slope of one, zero intercept, and zero noise variance. This hypothetical system is able to follow the target velocity exactly (unity gain, no bias) and with perfect reproducibility across trials (no variance). The parameter that is of interest in this study is the one that quantifies the variability of the real smooth pursuit system. In the linear model shown in Figure 3A, the Root Mean Square Error (RMSE) captures the variability of smooth pursuit.

In order to observe the effect of target size on the pursuit variability, we have plotted (Figures 3B) the estimated RMSE of the linear fits for the three target sizes in a representative subject (the same subject as in Figure 3A). As can be seen in the figure, increasing the diameter of the patch of random dots from 2° to 6° decreases the variability of pursuit initiation (Cohen’s d effect size = 1.59, Wilcoxon signed rank test: z = 26.92, p < 0.0001), while increasing the patch size from 6° to 20° increases the variability (Cohen’s d effect size = 1.89, Wilcoxon signed rank test: z = 27.39, p < 0.0001). Thus for this subject the relationship between eye velocity and stimulus size is consistent with the idea that pursuit initiation is driven by the population of surround-suppressed MT neurons (Figure 2B).

We next examined the relationship between pursuit variability and stimulus size for our cohort of seven subjects. To summarize the data from all subjects, we fitted linear models to all the data we acquired from individual subjects (see Methods for the details). The lines shown in Figure 4A are the averages of the estimated models across the subjects for the three conditions. For a simpler presentation of the data, and to make the plot less cluttered by individual data points, we grouped single pairs of eye/target velocity to quantiles. Each circle in Figure 4A represents one of these estimated target/eye velocity quantiles. The differences in the RMSE of the fitted models, as can be seen in this figure, capture the effect of stimulus size on smooth pursuit variability. The changes in the average pursuit variability across subjects is plotted in Figure 4B. The estimated RMSE values are normalized for each subject before averaging. The error bars are the standard error across subjects. The results indicate that, although there was substantial variability, the multi-subject analysis yielded effects similar to those shown for the example subject in Figure 3 (see Figure S1 for the estimated RMSE for the individual subjects). For the pursuit variability, we see a decreased RMSE from 2° to 6° (Cohen’s d effect size = 0.4376, Wilcoxon signed rank test: z = 1.5978, p = 0.1101), and a large increase from 6° to 20° (Cohen’s d effect size = 1.3112, Wilcoxon signed rank test: z = 2.0863, p = 0.0372). As reported previously (Born et al., 2000), there was also a decrease in pursuit gain when stimulus size increased beyond 6 degree (Cohen’s d effect size = 1.7473, Wilcoxon signed rank test: z = 2.5893, p = 0.0096).

In general, the results presented in this section showed that increasing the size of the pursuit target can decrease and then increase the pursuit initiation variability. As briefly explained in the Introduction, this pattern of results could arise from the sensitivity of the neurons involved in the sensory representation of...
target motion. To explore this point in greater detail, we performed simulations of a detailed computational model based on the properties of MT neurons.

**Computational model**

To suggest a plausible neuronal mechanism for the behavioral effects reported in the previous section, we devised a model that combines well-established aspects of motion coding (Hohl et al., 2013) with recent data on noise correlations in MT (Liu et al., 2016). We then tested different versions of this model on the same task as in our psychophysics experiments.

As detailed in the Methods, the model has two main stages. The first transforms the stimulus into a pattern of activation across a simulated population of MT neurons, with standard functions representing tuning for direction, speed, and size. We also included noise correlations among neurons and the dependency of these correlations on other tuning parameters (Liu et al., 2016). This provided an estimate of the neuronal response across MT for each stimulus used in our experiments.

The second stage involves a readout of the MT population, in which vector averaging is used to map MT activity into smooth pursuit velocity (Lisberger, 2010). Our formulation of the vector averaging operation provided the possibility of assigning different weights to different neuron types (surround-suppressed versus non-suppressed), which allowed us to consider different hypotheses about the readout of the MT population. In principle, the readout could make use of the entire MT population, as is often assumed. Or it could rely primarily on the MT subpopulation with strong surround suppression, as suggested by the results of Born et al. (2000). To evaluate these possibilities, we fit the model weights to the behavioral data reported in the previous section. The parameters of the simulated MT population (e.g. noise correlation structure) were chosen to maximize the similarity of the simulated and real MT population (see the Methods for details).

In the model, the vector averaging operation involves parameters corresponding to the integration area in visual space and the weights assigned to each neuron depending on its suppression index. The integration area limits the vector averaging only to the neurons with receptive field centered within a specific distance \(D\) from the stimulus center, as suggested by previous data (Mukherjee et al., 2016). The suppression index-related parameter \(a\) determines the contribution of each neuron in the vector averaging based on its surround suppression index.

Manipulation of these two parameters led to a range of models that varied in terms of the behavioral surround suppression (i.e. the ratio of the variability at 20 degrees to the variability at 6 degrees of stimulus diameter; Equation (18)). As detailed in the Methods, we used an optimization procedure to select the values of \(D\) and \(a\) (Figure 5). In short, we evaluated the behavioral surround suppression index of our model (Equation 18) for different values of \(D\) and \(a\) on a two-dimensional grid, and the parameter values that yielded the closest fit to our experimental data were selected. The optimal parameters were \(D = 4°\) and \(a = 15\), which correspond to a modest integration radius (Liu et al., 2016; Mukherjee et al., 2017) and a strong reliance on surround-suppressed MT neurons (Born et al., 2000).

**Figure 5.** The variability of smooth pursuit initiation simulated with our model. The trial-to-trial variability of the simulated pursuit velocity for different stimulus sizes of 0.4–20.4 degrees. The red curve illustrates the variability for the simulations based on \(a = 15\) (optimal value; only surround suppressed neurons contribute in the readout). The blue curve shows the simulations corresponding to \(a = 0\) (equal contribution of all neuron types in the readout). Both simulations are based on \(D = 4\) (optimal value). The error bars are calculated by iterating the simulations 1000 times.
Once the parameters were established, we calculated the variance of the model output by setting the pursuit target motion parameters (direction and velocity) to specific values, and iterating the simulations 1000 times. In the model, variability arises from the noise distributions assigned to each neuron and the noise correlation structure across the population (see Methods, Equation 7–Equation 9). Figure 5 shows the variance of the model output as a function of stimulus size for the optimal parameter values (red line); for comparison we also show the model output for the case in which equal weight is assigned to all neurons, regardless of surround suppression strength ($\alpha = 0$; blue line).

As can be seen in Figure 5 (red curve), increasing the size of the target initially (from $0.4^\circ$ to $4.4^\circ$), decreases the trial-to-trial variability of the decoded velocity, while further increases in stimulus size increase the variability. This pattern is similar to our psychophysical observations. In contrast, when surround-suppressed MT neurons are not accorded greater weight (Figure 5, blue line), the decoding velocity shows little dependence on size beyond a certain point.

Two factors can explain the initial increase in the precision of smooth pursuit, which was observed in both our psychophysics and simulated data. Relative to the smallest stimuli, larger stimuli induce a larger population response, both by filling the classical receptive fields of the neurons and by activating more neurons; the same two factors improve the signal-to-noise ratio (SNR) of the MT population (Figure 6). The more interesting phenomenon is the decrease in pursuit precision with further increases in stimulus size. In this case, a model that gives equal weight to all MT neurons predicts a saturation in pursuit precision with increasing stimulus size (Figure 5, blue lines), while a model that gives greater weight to surround-suppressed neurons predicts the size-dependent effects we have seen in our data (Figure 5, red lines). The decreased precision in the model arises from the decreased sensitivity of surround-suppressed MT neurons for increasing stimulus size. This decreased sensitivity for large stimuli is evident in the average firing rate of the surround-suppressed MT neurons (Figure 6, left, red line) compared to that of the non-surround-suppressed neurons (Figure 6, left, blue line).

**Discussion**

Electrophysiological recordings have identified a range of surround suppression strengths among MT neurons (Allman et al., 1985). Neurons with strong surround suppression are clustered topographically within MT (Born & Tootell, 1992) and form separate projections to higher-level cortex (Berezovskii & Born, 2000). Furthermore, MT neurons with surround suppression are more activated for stimuli of short duration, as necessarily occurs during the onset of pursuit, while non-surround-suppressed neurons respond poorly to brief stimuli (Churan et al., 2008; Tsui & Pack, 2011). These considerations suggest that surround-suppressed neurons have a particularly important influence on the initiation of smooth pursuit (Born et al., 2000). Through psychophysics experiments and computational modeling, we have provided additional evidence in this paper that MT neurons with surround suppression are likely to be the main contributors to smooth pursuit initiation.

---

**Figure 6.** The average firing rate (left) and signal-to-noise ratio (SNR, right) for the population of simulated MT neurons. The insets are the enlarged version of the marked parts in the main graphs. The circled points A and B correspond to the firing rate (on left) and SNR (on right) at $2.2^\circ$ and $4.4^\circ$ of stimulus size, respectively. The red and blue curves correspond to the neurons with and without surround suppression.
Furthermore, we modified previously proposed models of pursuit initiation (Hohl et al., 2013; Huang & Lisberger, 2009) to suggest a neuronal mechanism for our behavioral observations. Unlike previous models, our proposed model encodes motion parameters in a heterogeneous population of MT neurons (surround suppressed and non-surround suppressed) which are different in terms of their stimulus sensitivity and their structure of noise correlation (Liu et al., 2016). Our model is sufficiently accurate in reproducing the structure of noise correlation recorded from area MT (Figure 5), and is able to accurately replicate the pattern of behavioral variability observed in our psychophysics experiments (Figure 6). Although the physiological differences between surround suppressed and non-surround suppressed neurons had been shown previously (Allman et al., 1985; Liu et al., 2016) and their distinct roles in smooth pursuit had been suggested (Born et al., 2000; Komatsu & Wurtz, 1988), for the first time in this paper, we proposed a computational model that yields a reasonable simulation of smooth pursuit initiation taking into account the differential contributions of these neurons to behavior.

Sources of motor variability
Our results have important implications for understanding the neuronal basis of behavioral variability in smooth pursuit initiation (Osborne et al., 2005). Previous studies attributed a large portion of this variability to the noisy responses of MT neurons. Specifically, by showing that noise correlations in area MT exhibit a specific structure that is detrimental to motion coding, Huang & Lisberger (2009) claimed that the neuronal trial-to-trial variability propagates from area MT all the way through to behavior. This hypothesis, of a sensory source for motor variability, is based on the assumption that area MT is a homogenous population of neurons with identical noise correlation structure and equal contributions to the encoding of stimulus motion during smooth pursuit initiation. However, Liu and colleagues showed recently that the noise correlation between MT neurons with surround suppression is less detrimental for motion coding than other MT subpopulations (Liu et al., 2016). Specifically, because the noise correlations are largely unrelated to the signal correlations, the noise in the responses can be averaged out, to a large extent, over the encoding population. Hence the variability in the output of MT could be considerably lower than previously suspected, such that the limiting factors for pursuit precision are the sensitivity of the MT population and the noise introduced as part of the sensorimotor transformation. This interpretation leaves open the possibility that other areas downstream of area MT also contribute in the variability of smooth pursuit initiation (Lee et al., 2016).

Surround suppression in the gain of smooth pursuit
Previous studies have largely focused on the gain of smooth pursuit and its relationship to stimulus parameters. For example, low-contrast stimuli shift the population of MT neurons toward an overestimation of stimulus speed, as neurons that encode high speeds become active for slow stimuli (Pack et al., 2005). At the same time, pursuit gain actually decreases with decreasing contrast (Priebe & Lisberger, 2004). To reconcile these results, Priebe & Lisberger (2004) made use of a modified version of vector averaging, in which an additional term was introduced in the denominator to decrease the estimated velocity when the average firing rate in MT is low.

This approach can be accommodated within the framework of Bayesian motion coding (Weiss et al., 2002), wherein, low contrasts are associated with less precise stimulus coding, which leads to a greater reliance on prior assumptions. In models that include a prior belief that objects are stationary (i.e. zero mean velocity prior), decreasing the precision of image motion leads to a lower perceived velocity; this effect has been confirmed psychophysically (Ascher & Grzywacz, 2000; Hürlimann et al., 2002; Stocker & Simoncelli, 2006).

The relationship between the current results and the Bayesian coding scheme is not entirely clear, as the larger stimulus can in principle be integrated so as to reduce the precision of the velocity estimate. However, it does not appear that the brain performs this integration, given the observation that pursuit gain decreases for large stimuli (Born et al., 2000), while the variability increases (Figure 4). These findings highlight the importance of considering the neural representation of the stimulus, which in this case appears to be determined primarily by surround-suppressed MT neurons.

It should also be noted that the Bayesian model depends heavily on the form of the prior assumption, and it is not difficult to find experimental situations in which the low-velocity prior is violated (Thompson et al., 2006). Thus, the relationship between MT activity and the gain of pursuit initiation is likely more complex than what has been captured by existing models (Brooks et al., 2011). We have largely avoided this issue by focusing on the variability of pursuit initiation.

Relation to oculor following response
A similar behavior to smooth pursuit eye movement is the ocular following response (OFR), which is a short-latency involuntary eye movement elicited by sudden movement of large-field visual stimuli (Gellman et al., 1990; Miles et al., 1986). These movements are important for stabilizing the eyes on the visual scene. The OFR and smooth pursuit eye movements share part of their neuronal pathways, including brain areas MT and MST, as well as other subcortical regions (e.g. cerebellar nuclei) (Kawano et al., 1994; Optican et al., 1986). Given their commonalities, one might suspect that the OFR system partially interferes with the smooth pursuit system in our psychophysics experiments, and could be responsible for at least part of our behavioral effects. Particularly, given the fact that OFR is triggered by large-field moving stimuli, the largest stimulus in our psychophysics task (20 degrees) could potentially engage the OFR system. However, certain aspects of our results challenge this possibility. First: as mentioned in the Results and reported elsewhere (Born et al., 2000), the gain of eye movements in our task decreases for the largest stimulus compared to the smaller one. This contradicts our expectations of OFR gain which should increase monotonically with stimulus size (Masson & Perrinet, 2012; Miles et al., 1986). Second, in our paradigm, subjects are instructed to follow a stimulus whose motion is consciously anticipated. Unlike our task, in conditions where OFR has been reported, the subjects were unaware of stimulus motion and their eye movements were
involuntary (Glasser & Tadin, 2014). Third, our moving texture (random dots) appears on the screen at the same time that the main target starts to move in the periphery (see Experimental Procedure). In the OFR paradigms, however, the stationary texture on the screen starts drifting at an unexpected time. Moreover, the absence of foveal motion signal decreases the OFR gain significantly (Miles et al., 1986). Together, these aspects of our task paradigm and behavioral results reduce the chance of OFR system involvement in our observations, albeit do not entirely dismiss its possibility.

In summary, we showed that the neurons with surround suppression in area MT are the main contributors to smooth pursuit initiation. This causes a specific pattern of changes in the trial-to-trial variability of pursuit initiation that was validated by our psychophysics experiments. In addition, our proposed model was shown capable of reproducing this behavioral pattern.

Data and software availability
All the human psychophysics data and analysis codes are available on GitHub and are archived in http://doi.org/10.5281/zenodo.1469149 (Bakhtiari, 2018).

Supplementary material
Supplementary Figure 1. The estimated RMSE values for all 7 subjects for the three stimulus sizes. Gray lines represent the data from each subject.

References


Open Peer Review

Current Peer Review Status: ✓ ✓

Version 2

Reviewer Report 25 April 2019

https://doi.org/10.21956/mniopenres.13880.r26155

© 2019 Blohm G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ Gunnar Blohm
Centre for Neuroscience Studies (CNS), Queen's University, Kingston, ON, Canada

The authors did a great job addressing my questions and concerns and I have no further comments. Thanks a lot for all your clarifications.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 18 April 2019

https://doi.org/10.21956/mniopenres.13880.r26154

© 2019 Price N. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ Nicholas S. C. Price
Department of Physiology, Monash Biomedicine Discovery Institute, Australian Research Council Centre of Excellence for Integrative Brain Function, Monash University, Clayton, Vic, Australia

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
This is an interesting combination of electrophysiology, modeling and psychophysics that addresses the question of how surround suppression affects the variability in the velocity of pursuit initiation. The modeling and decoding of the physiological data are well developed, but the results of the psychophysics are a little difficult to follow and the associated presentation is confusing. In order to trust that the dip in RMSE at intermediate stimulus sizes is a robust finding, I would suggest that it needs to be demonstrated within the majority of the individual observers. Suggestions for how to do this are below. Once the psychophysical result is more robustly demonstrated, I think it will be appropriate to compare it to the modeling results.  

**Major**

1. In Fig. 3, data for a single subject is shown. Error bars show standard error, but what is the sample here? Isn’t this based on linear regression to a single subject’s data? Why is there more than a single value?

2. In Fig. 4, the marginal values are not helpful, and it’s not clear why the data has been grouped into quantiles (quintiles?) Possibly, this is because data from all subjects is shown and otherwise there are too many data points, but this could be explained. Fig 4A (left) shows data pooled across all subjects, whereas Fig 4B (right) shows data analysed for individual subjects, before averaging. The latter approach seems more appropriate, however, it would be improved by showing raw data for individual subjects. Are RMSEs in Fig 4 normalized? How was this performed? If no normalisation was performed, how can the scale of the RMSEs (0-1) be reconciled with the values in Fig. 3 (1.6-2.2)? Why not show raw data for individual subjects before normalising? This would be far more convincing in terms of demonstrating that the dip at 6deg occurs in most or all subjects.

3. I appreciate that the authors have chosen to focus on variability, hence they use RMSE. However, their stimulus conditions also affect the gain of pursuit, and it’s not clear how this interacts with the RMSE. This might be discussed, for instance at the top of p10, when it is stated that RMSE captures the variability of smooth pursuit. In the statistical analysis associated with Fig 4, it would be helpful if individual, rather than pooled, data were analysed.

4. Page 11 states “We used an optimization procedure to select the values of D and a”. What was this procedure? What exactly was being optimized?

**Minor**

1. Methods – Delete sentence “The experiments were carried out in total darkness.” There was a visual stimulus, therefore, the room cannot have been completely dark.
2. Methods – Target luminances are provided; what was the white dot luminance?
3. How were the 3 sizes selected?
4. Figure 1. Consider plotting eye speed as well as, or in addition to, eye position. The x-axes could be more clearly labeled (e.g. Time since motion onset / Time since pursuit onset). More importantly, it’s not clear how helpful these graphs are, since they show a range of stimulus conditions, and alignment in the lower panel is 20 ms after the saccade. Given the presence of corrective saccades, which aspect of the saccade is used for alignment? Given that eye acceleration occurs during initial pursuit, have the authors considered whether the choice of the time window (20-80 ms) affect the results in any way?
5. Figure 2b. Consider adding a legend so that it is clear what triangles/circles and red/blue indicate (I realise this is in the text, but it would be helpful in the figure). It’s not clear what the spatial structure of the boxes marked “Area MT” and the dotted line are meant to represent, nor how the colours and markers showing neurons within the boxes relate to the tuning curves shown at above the second column. “… depicted in the fourth _column_” (not _row_).
6. Figure 4. Add labels A/B. Add “n=7” to Fig 4B (if appropriate).
7. Figure 6. X-label is misspelled “diameter”.
8. Discussion (p12), it is stated that “MT neurons are more activated for stimuli of short duration”. Is this referring to rapid adaptation? It’s not clear why a neuron should respond more or less, before it “knows” how long a stimulus will continue for.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Visual motion processing; oculomotor system; modeling

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 04 Mar 2019
Shahab Bakhtiari, McGill University, Montreal, Canada
1) In Fig. 3, data for a single subject is shown. Error bars show standard error, but what is the sample here? Isn’t this based on linear regression to a single subject’s data? Why is there more than a single value?

We used bootstrapping to generate many samples for the estimated variability. We used the `bootstrp()` function in MATLAB to generate 100 estimates of variability per stimulus size:

```matlab
>> allRMSE = bootstrp(100,@estimateRMSE,targetV,EyeV);
```

```matlab
function rmse = estimateRMSE(TargetV,EyeV)
    mdl = LinearModel.fit(TargetV,EyeV);
    rmse = mdl.RMSE;
end
```

The error bars in Figure 3 show the standard error calculated on the bootstrapped samples of RMSE. In the revised version, we added a short explanation about the bootstrapping in the figure caption.

2) In Fig. 4, the marginal values are not helpful, and it’s not clear why the data has been grouped into quantiles (quintiles?) Possibly, this is because data from all subjects is shown and otherwise there are too many data points, but this could be explained. Fig 4A (left) seems to show data pooled across all subjects, whereas Fig 4B (right) shows data analysed for individual subjects, before averaging. The latter approach seems more appropriate, however, it would be improved by showing raw data for individual subjects. Are RMSEs in Fig 4 normalized? How was this performed? If no normalisation was performed, how can the scale of the RMSEs (0-1) be reconciled with the values in Fig. 3 (1.6-2.2)? Why not show raw data for individual subjects before normalising? This would be far more convincing in terms of demonstrating that the dip at 6deg occurs in most or all subjects.

This comment includes several points that we address one by one below:

**Figure 4A:** As you mentioned, in Figure 4A, data points are grouped into quantiles to reduce the complexity of the figure due to too many samples (pooled data across 7 subjects). Showing the pooled version of the data helps understanding the similar range of eye velocities across subjects. Also, the marginal histogram demonstrates the range of values of target and eye velocities for different conditions. We added a short segment to the Results section to clarify the presentation of the data in Figure 4A.

**Figure 4B:** Figure 4B shows the results of fitting individual linear models to each subject after averaging the estimated normalized RMSE values. The error bars show standard error across subjects. The alternative approach, as explained in the Methods section, is to pool the data from all subjects and fit a mixed-effect linear model. As mentioned in the Methods, both approaches give the exact same results. In the revised version of the paper, we’ve added a supplementary figure that shows the results of individual fits for all the subjects (Figure S1). It is also added in the Results section of the revised manuscript that RMSE values of individual subjects can be found in the supplementary figures. However, due to the relatively large inter-individual differences in the baseline variability across subjects (see Figure S1), we decided to keep the normalized RMSE
values and the average across subjects in Figure 4B. It's more straightforward to see the main effect in the average data, while not getting confused by the inter-individual differences across subjects. Given the consistency of the main effect across subjects, averaging across subjects does not lead to any misrepresentation of the data.

Another important point in this comment concerns the variability of pursuit initiation at the 6-degree stimulus (medium size stimulus). The dip at the 6-degree stimulus size happens in 5 subjects (out of 7). Our statistics show that the difference between the variabilities at 2-degree and 6-degree stimuli is not statistically significant across subjects (p=0.11) probably because even in those subjects that show this effect, the effect size is small (see Figure S1). Sampling the variability along the stimulus size axis with higher resolution could have given us a more realistic estimation of the stimulus size at which the minimum variability happened for each subject. However, due to the time-consuming nature of the experiment, we could only sample the variability for three stimulus sizes. Although important questions could be asked regarding the stimulus size that induces the lowest variability, the main hypothesis of the current study concerned the variability for the largest stimulus size. The largest trail-to-trial variability happens in the largest stimulus size in 5 out of 7 subjects, and in these 5 subjects, the effect size is quite large, and the effect is statistically significant across subjects. In two of the subjects, increasing the stimulus size does not change the variability (in one subject), or decreases it (in the other one). This difference between subjects in terms of spatial integration is expected and have been reported elsewhere (Tadin et al. 2015, Vision Research).

3) I appreciate that the authors have chosen to focus on variability, hence they use RMSE. However, their stimulus conditions also affect the gain of pursuit, and it's not clear how this interacts with the RMSE. This might be discussed, for instance at the top of p10, when it is stated that RMSE captures the variability of smooth pursuit. In the statistical analysis associated with Fig 4, it would be helpful if individual, rather than pooled, data were analysed.

We agree that the relationship between pursuit gain and variability is interesting to investigate. In our data, as we mentioned in the Results section, in addition to the changes in pursuit variability, the pursuit gain also changes as a function of size (subsection Psychophysics experiments, at the end of the 6th paragraph). Particularly, increasing the size of the stimulus decreases the gain of smooth pursuit (in 6 out of 7 subjects - Cohen's d effect size = 1.7473, Wilcoxon signed ranked test: z = 2.59, p = 0.0096). In the Discussion section (subsection Surround suppression in the gain of smooth pursuit), we discussed this observation in the context of previous studies, and how the decrease in the gain might be related to the increase in pursuit variability for large stimuli. In short, previous studies have shown that in case of high uncertainty (low contrast stimuli, etc.), the gain of pursuit decreases. As pursuit uncertainty (or variability) increases for large stimuli in our psychophysics data, according to previous studies, we expect the decrease in the gain of pursuit for large stimuli. However, since gain of smooth pursuit was not directly related to the hypothesis that we wanted to test in this study, we did not go deeper on the causes and consequences of pursuit gain modulation with stimulus size.

4) Page 11 states “We used an optimization procedure to select the values of D and a”. What was this procedure? What exactly was being optimized?

This is an important point, and we tried to clarify it in the revised manuscript. For the two parameters $a$ and $D$, which determine the size of spatial integration and the contribution of MT neurons in the pursuit initiation respectively, we ran a two-dimensional grid search. For every pair
of (a,D) values, we evaluated the behavioral suppression index (equation 18) of our model output; namely, we measured the ratio of variability at 20 degrees and 6 degrees stimulus sizes. The pair that gave us the most similar suppression index to our averaged behavioral data (a=15, D=4) was chosen for the final simulation results. This search optimization procedure is explained in the Methods (subsection Decoding stimulus velocity) and also briefly explained in the Results section of the revised manuscript.

Minor
1) Methods – Delete sentence “The experiments were carried out in total darkness.” There was a visual stimulus, therefore, the room cannot have been completely dark.

We deleted that sentence.

2) Methods – Target luminances are provided; what was the white dot luminance?

We added the information about the dots luminance in the Methods.

3) How were the 3 sizes selected?

We chose the three stimulus sizes in a way that evokes three levels of MT responses: weak response when the stimulus is smaller than the receptive field (2 degrees), strong response when the stimulus fits the receptive field (6 degrees), and weak or strong response depending on the structure of the receptive fields for the two types of MT neurons (20 degrees). We chose these three stimulus sizes based on the population of real MT neurons that had been recorded in monkey electrophysiology experiments. You can see the average firing rate of our MT population for different stimulus sizes in Figure 6 (left), and you can notice that our selected stimulus sizes do trigger three levels of MT responses. Although we tried to target these three states of MT activity in our psychophysics experiments, there is an inter-individual difference across subjects which implies that the three selected sizes might be suboptimal in some of the subjects (especially the 6-degree stimulus). We think this might partly explain the inter-individual variability we see in our data.

4) Figure 1. Consider plotting eye speed as well as, or in addition to, eye position. The x-axes could be more clearly labeled (e.g. Time since motion onset / Time since pursuit onset). More importantly, it’s not clear how helpful these graphs are, since they show a range of stimulus conditions, and alignment in the lower panel is 20 ms after the saccade. Given the presence of corrective saccades, which aspect of the saccade is used for alignment? Given that eye acceleration occurs during initial pursuit, have the authors considered whether the choice of the time window (20-80 ms) affect the results in any way?

The horizontal axis of the plots in Figure 1 is modified to address your concern.
Also, in the revised manuscript, the post-saccadic eye velocity is added to Figure 1. As we explained in response to Dr. Blohm’s comment on the selected time window, there were some considerations in choosing 20-80 ms after saccade offset for further analysis:
First, after saccades offset, there is an acceleration phase (about 20 ms) that we tried to exclude it from our analysis time window.
Second, we made sure that within this time window, for all our subjects, there is relatively constant eye velocity. We decided to use the same time window for all subjects to avoid any further
variability across subjects due to the arbitrary selection of the time window for each subject separately. Third, since our eye velocity measurements are noisy, we decided to choose a time window for the analysis, instead of one single time point, and average the eye velocity within this time window as a representation of eye velocity for each trial. We this approach, we made sure that the measurement noise does not confound the behavioral variability. Fourth, to make sure that we didn’t include the close-loop phase of smooth pursuit in our analysis, we chose the end of the time window at 80 ms which is about 20 ms less than the well-known 100-ms duration of open-loop pursuit.

5) Figure 2b. Consider adding a legend so that it is clear what triangles/circles and red/blue indicate (I realise this is in the text, but it would be helpful in the figure). It’s not clear what the spatial structure of the boxes marked “Area MT” and the dotted line are meant to represent, nor how the colours and markers showing neurons within the boxes relate to the tuning curves shown at above the second column. “… depicted in the fourth _column_” (not _row_).

Thanks for the suggestions. We added a legend to the figure that explains the meaning of triangles/circles and red/clue color code. Also, few sentences are added to the figure caption that clarify the spatial structure of the schematic area MTs, and how the color of circles/triangles are related to the size tuning curves in the first row.

6) Figure 4. Add labels A/B. Add “n=7” to Fig 4B (if appropriate).

Figure 4 is modified based on your suggestion.

7) Figure 6. X-label is misspelled “diameter”.

Corrected! Thanks.

8) Discussion (p12), it is stated that “MT neurons are more activated for stimuli of short duration”. Is this referring to rapid adaptation? It’s not clear why a neuron should respond more or less, before it “knows” how long a stimulus will continue for.

Thank you for highlighting this sentence. There was an error in this sentence that needed to be corrected. The sentence is now corrected to: “MT neurons with surround suppression are more activated for stimuli of short duration.”

A previous study from our lab (Churan et al., 2008, Current Biology) showed that MT neurons with surround suppression integrate motion stimuli more quickly compared to non-surround-suppressed neurons. Therefore, if we measure the responses of MT neurons to a stimulus with fixed short duration a larger response is evoked in surround-suppressed neurons compared to non-surround suppressed neurons.

**Competing Interests:** No conflict of interest.
This manuscript investigates how different functional sub-populations of MT neurons might be involved in smooth pursuit initiation. The combination of behaviour and physiologically constrained modelling allows identifying surround-suppressed neurons to be the main drivers of pursuit initiation.

The approach is sound and well carried out. I do have a few questions relating clarity and assumptions that I think would further improve this study.

1) It is not quite clear to me why the authors suggest that additional noise sources (presumably downstream of MT) also contribute to movement variability. I might have missed the logic here but maybe the authors could lay the reasoning out more clearly early on the manuscript?

2) Eye movement analysis. Your measurement window for pursuit initiation relies on a tight coupling between saccade execution and pursuit initiation. Both can have quite some inherent variability. It would be good to validate this hypothesis by explicitly plotting trial-by-trial pursuit vs saccade latencies in a scatter plot. This would in part validate Figure 3.

3) Why was average velocity used? Was it constant? You expect pursuit initiation to be accompanied by eye acceleration. It would be nice to plot eye velocity in addition to position in Figure 1 to convince the reader that this was appropriate. This is also in relation to Figure 3.

4) Statistical analysis mentions fitting methods. Could you please specify which ones were used and why?

5) Decoding stimulus velocity. why was vector averaging used instead of ideal observer / Bayesian optimal decoding? Would your results hold if other methods were applied? It is not clear to me that other methods would lead to the same conclusions.

6) On page 11, top right: you mention an optimization procedure but I couldn't find any details. What was used and why?

7) The main difference between your model and the pursuit data is that variability for very small diameter stimuli in the data is lower than variability for very large stimuli (the opposite was the case in your model). Is that a consistent effect in the data? I.e. significant? Why do you think the model predicts something different?

8) A detail: last paragraph on page 12, I'm not sure you refer to the correct figures. Otherwise I did not understand this paragraph...

Is the work clearly and accurately presented and does it cite the current literature?
Yes
Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

**Author Response 04 Mar 2019**

**Shahab Bakhtiari**, McGill University, Montreal, Canada

1) It is not quite clear to me why the authors suggest that additional noise sources (presumably downstream of MT) also contribute to movement variability. I might have missed the logic here but maybe the authors could lay the reasoning out more clearly early on the manuscript?

We agree that the link between our results and the hypothesis of non-sensory sources of pursuit variability was not sufficiently clear in the Introduction. In the revised Introduction section (third paragraph), we now added a part that states the difference in variability between the two subtypes of MT neurons, and tried to make it clear that this difference can have important implications for pursuit-MT relationship, especially if the two subtypes do not contribute equally in pursuit initiation. In short, previous studies claimed that MT variability underlies pursuit variability assuming that the co-variability across MT population has a structure that cannot be averaged out by the decoder (Huang, Xin, and Stephen G. Lisberger, 2009, *Journal of Neurophysiology*). These studies also assumed that area MT is a homogenous population with all the neurons equally contributing in pursuit initiation, and all having the same structured co-variability (so called short-range noise correlation). Based on our psychophysics data and computational modeling, we suggested that only a subset of neurons (neurons with surround suppression) mainly contribute in pursuit initiation. A previous study from our lab (Liu, Liu D., Ralf M. Haefner, and Christopher C. Pack, 2016, *eLife*) showed that this subpopulation of MT exhibits a noise correlation that is not detrimental to stimulus coding, and can be averaged out by the decoder. Based on our results, the subpopulation of MT that drives the smooth pursuit initiation is not that noisy at the end, which implies that the variability in pursuit initiation should have sources other than MT neurons.

2) Eye movement analysis. Your measurement window for pursuit initiation relies on a tight coupling between saccade execution and pursuit initiation. Both can have quite some inherent...
variability. It would be good to validate this hypothesis by explicitly plotting trial-by-trial pursuit vs saccade latencies in a scatter plot. This would in part validate Figure 3.

This is an important point indeed. We agree that there is, in general, some inherent variability in saccade initiation. Since the time window that we selected for our further analysis was coupled to the saccades offset (20-80 ms after saccade offset), the saccade variability could potentially affect our pursuit analysis. For example, saccade onset could have happened after or before the initiation of pursuit, and the selected time window (20-80 ms after saccade offset) could have been in different phases of smooth pursuit. However, the variability in saccades timing would be a problem in our approach if there were pre-saccadic pursuit. In our behavioral data, pursuit initiation always happens after the saccades. In our data, there is no significant non-zero eye velocity prior to the saccade initiation (t-test on the average eye velocities in 0-20 ms before the saccade, p=0.67), and about 10 ms after the saccade the eye velocity becomes significantly larger than zero (t-test on the average eye velocities in 10-30 ms after the saccade, p<1e\(^{-5}\)). Therefore, pursuit initiation happens after saccade offset, regardless of the inherent variability of saccade timing.

3) Why was average velocity used? Was it constant? You expect pursuit initiation to be accompanied by eye acceleration. It would be nice to plot eye velocity in addition to position in Figure 1 to convince the reader that this was appropriate. This is also in relation to Figure 3.

The estimated eye velocities were noisy for individual trials in our data. Ideally, we were interested in true trial-to-trial variability of the eye velocity that is not confounded with measurement noise. To minimize the impact of measurement noise on the estimated trial-to-trial variability in pursuit initiation, and improve the power of our estimates, we averaged the eye velocity for each single trial in a time window (20-80 ms).

Regarding the eye acceleration within this time window, there is a small increase in eye velocity from the beginning to the end of this time window, but the main eye acceleration happens from 0-20 ms (in the revised version of Figure 2, you can see the eye velocity for this time window). The main increase in the average eye velocity happens from 0 to 20 ms after the saccade. We can still see an acceleration from 20 ms to 80 ms but it’s smaller in amplitude. The 20-80 ms period is chosen in a way that for all subjects, it gives a reasonable representation of eye velocity for pursuit initiation, and does not get too close to the end of the open-loop pursuit initiation phase (~100 ms).

4) Statistical analysis mentions fitting methods. Could you please specify which ones were used and why?

To model the relationship between the eye and target velocities, we could either fit one linear regression model per subject (equation 1), or pool all the data from all the subjects and fit a mixed-effect model with subjects’ identities as the random-effect (equation 2). We’ve made it clear in the revised version (in Statistical analysis section) that, first, our final conclusion holds regardless of which approach was taken for the linear model, and second, the results reported in figures 3 and 4 are based on the one-model-per-subject approach (equation 1).

Also, in the revised version, we’ve included a supplementary figure that shows the estimated RMSE values for the three condition for each subject individually before averaging (the average was shown in Figure 4).

5) Decoding stimulus velocity. why was vector averaging used instead of ideal observer / Bayesian optimal decoding? Would your results hold if other methods were applied? It is not clear to me that
other methods would lead to the same conclusions.

We used vector averaging as the decoder because previous studies have shown that, compared to other types of decoders, vector averaging (particularly the variant that we used in the current study) better predicts different aspects of pursuit initiation. Particularly, previous studies from Lisberger’s lab (Hohl, et al. 2013, Neuron) compared different variants of vector averaging, as well as the ideal observer, as the decoder in their pursuit initiation model, and showed that the relation between MT responses and pursuit parameters can be best captured by the same version of vector averaging that we used in this study.

That being said, we can also show that our final conclusions should hold if we use the ideal observer as the decoder. The accuracy of the optimal linear decoder depends on the information represented by the population which can be measured by Fisher Information (Moreno-Bote, et al. 2014, Nature Neuroscience):

\[ I = f'\Sigma f', \]

where \( f \) represents the tuning curve of the neurons, and \( \Sigma \) is the population covariance. Also, for discriminating between two stimulus values, \( s^+ \) and \( s^- \) (e.g. two target velocities):

\[ f' = f(s^+) - f(s^-). \]

Based on the Fisher information equation, the information represented by a sensory population depends on the difference in the firing rate of the neurons for the two stimulus values, and the structure of the covariance matrix.

In smooth pursuit, increase in pursuit variability (or noise) corresponds to decrease in Fisher information of the population. The lower the information represented by the sensory population, the higher the variability of the decoded target velocity. For the surround-suppressed neurons in area MT, decreases for large stimuli as the overall sensitivity of the neurons decrease (\( f' = f(s^+) - f(s^-) \rightarrow 0 \)). Decrease in \( f' \) results in decrease of Fisher information, based on the equation above, and increase in pursuit variability.

Therefore, even if we used an optimal decoder in our model, for large stimuli, the pursuit variability would increase when greater weights are assigned to the surround-suppressed neurons; the same observation that we had with the vector averaging.

6) On page 11, top right: you mention an optimization procedure but I couldn't find any details. What was used and why?

In the Methods section (Decoding stimulus velocity subsection), we explained that the values of the parameters \( a \) and \( D \) are set by a two-dimensional grid search on a range of values for the two parameters. The values that give the most similar behavioral surround suppression index to our psychophysics results are selected. To address your concern and better clarify the optimization process, we added a short description in the Results section of the revised manuscript.

7) The main difference between your model and the pursuit data is that variability for very small diameter stimuli in the data is lower than variability for very large stimuli (the opposite was the case in your model). Is that a consistent effect in the data? I.e. significant? Why do you think the model predicts something different?

The smallest stimulus size in the real data and the model are not the same. In the model, the smallest size is 0.5 degree, while in the behavioral data, it’s 2 degrees. Both the model and the behavioral data have similar ranges of variability close to the 2-degree stimulus size.

In our data, inter-individual variability across subjects makes it difficult to draw clear conclusions about the changes in the variability between the small and medium stimulus sizes. The trough of
the variability-stimulus size curve can potentially happen at slightly different stimulus sizes for different subjects. Due to time limitation, we could not sample the variability-size curve in a higher resolution (we sampled three stimulus sizes), and therefore, could not reliably detect the stimulus size in which the minimum behavioral variability happens. In the model, this uncertainty does not exist, as all the simulations are based on a fixed population of MT neurons for which the min variability happens at a specific stimulus size. That is, the variability-size curve is fixed for the MT population which is not the case in our cohort of human subjects. We expect that the min variability happens at slightly different stimulus sizes for every subject. Since evaluating variability for many stimulus sizes was very time consuming, we had to limit our experiment to three sizes which we predicted (based on the monkey MT data) to reveal the modulation in the variability that we expected to happen.

8) A detail: last paragraph on page 12, I'm not sure you refer to the correct figures. Otherwise I did not understand this paragraph...

In this paragraph, we explain that the decrease in pursuit precision (or increase in pursuit variability) for large stimulus sizes is due to the decreased sensitivity of the neurons with surround suppression. In Figure 6, we show that the average firing rate of the surround-suppressed neurons decreases when increasing the stimulus size while the non-surround suppressed neurons saturate for large stimuli. We observed that in our behavioral data, pursuit variability increased for large stimuli. In our model, pursuit variability increased for large stimuli only when larger readout weights are given to the neurons with surround suppression, which reflects the decreased sensitivity of surround-suppressed neurons for large stimuli.

We corrected some parts of this paragraph in the revised version (including the references to figures) and added a sentence to clarify.

**Competing Interests:** No conflict of interest.