



Neural mechanisms underlying the temporal control of sequential saccade planning in the frontal eye field

Debaleena Basu^{a,1}, Naveen Sendhilnathan^b, and Aditya Murthy^a

^aCentre for Neuroscience, Indian Institute of Science, Bengaluru 560012, India; and ^bDepartment of Neuroscience, Columbia University, New York, NY 10027

Edited by Robert H. Wurtz, NIH, Bethesda, MD, and approved August 20, 2021 (received for review May 13, 2021)

Sequences of saccadic eye movements are instrumental in navigating our visual environment. While neural activity has been shown to ramp up to a threshold before single saccades, the neural underpinnings of multiple saccades is unknown. To understand the neural control of saccade sequences, we recorded from the frontal eye field (FEF) of macaque monkeys while they performed a sequential saccade task. We show that the concurrent planning of two saccade plans brings forth processing bottlenecks, specifically by decreasing the growth rate and increasing the threshold of saccade-related ramping activity. The rate disruption affected both saccade plans, and a computational model, wherein activity related to the two saccade plans mutually and asymmetrically inhibited each other, predicted the behavioral and neural results observed experimentally. Borrowing from models in psychology, our results demonstrate a capacity-sharing mechanism of processing bottlenecks, wherein multiple saccade plans in a sequence compete for the processing capacity by the perturbation of the saccade-related ramping activity. Finally, we show that, in contrast to movement-related neurons, visual activity in FEF neurons is not affected by the presence of multiple saccade targets, indicating that, for perceptually simple tasks, inhibition within movement-related neurons mainly instantiates capacity sharing. Taken together, we show how psychology-inspired models of capacity sharing can be mapped onto neural responses to understand the control of rapid saccade sequences.

motor sequences | oculomotor control | electrophysiology | processing bottlenecks

Saccadic eye movements shift the fovea from one point to another, serially sampling our visual surroundings and aiding consequent behavior. Proper planning and execution of saccade sequences is essential for performing everyday tasks such as reading. Despite extensive research on the neural basis of planning individual saccades, the neural mechanisms underlying the sequencing of multiple saccades remain largely unknown. Previous research has shown that sequential saccades can be processed in parallel (1–15). Sequential saccade studies have shown that, as the temporal gap between the targets (target step delay; TSD) decreases, the latency of the response to the second stimulus increases markedly, as if the brain inherently cannot process two simple decisions at the same time (16, 17). The bottlenecks associated with parallel programming of multiple saccade plans form the basis of this study.

Various theoretical frameworks have been proposed to explain how closely spaced action plans interfere with each other. Single-channel bottleneck models propose that a central, decision-making stage constitutes the bottleneck, wherein the central stages of multiple plans can only proceed serially and cannot be “coactive” (16, 18–20). For a sequence of two saccades, the first plan is likely to reach the central stage first, and thus, the saccade 2 plan must “wait” until central processing of the first is over (Fig. 1A). In contrast, capacity-sharing models argue that the decision-making stages of both plans can proceed in parallel, albeit with differential rates. The concept of the brain’s “capacity” corresponds to the brain’s general information-processing capabilities (21–24), independent of task

type. The capacity-sharing models predict that, because of its temporal precedence, the first saccade plan will get the major share of the capacity, and the second saccade plan will get a smaller fraction, thus delaying the onset of the second response (Fig. 1B; 25, 26).

The neural mechanisms of processing bottlenecks in sequential saccade planning are not known. To investigate the neural architecture of saccade-related bottlenecks, we recorded from the frontal eye field (FEF) of macaque monkeys performing a sequential saccade task. FEF is a good candidate region to study the neural imprints of processing bottlenecks, since it is a higher-order control center for goal-directed saccadic planning (27–29). Furthermore, the activity of FEF movement neurons follow the dynamics of accumulator models and resemble the central, capacity-limited stage observed in computational models of dual-task studies (30–32). Finally, FEF movement neurons can encode two saccade plans in parallel (4), and thus, any limitations arising during the concurrent programming of saccades may be found in the activity of movement-related neurons in the FEF (4).

In this study, we show that FEF movement neurons constitute a bottleneck locus—the processing of saccadic sequences is slowed down by reducing the speed of activity growth or by increasing movement activation threshold. Such adjustments were observed for both the first and second saccade plans, indicating that a capacity-sharing mechanism might underlie temporal delays that limit the extent of parallel processing seen during the sequencing of multiple actions.

Significance

For accomplishing even the simplest of daily tasks, we execute rapid sequences of goal-directed saccadic eye movements that move from one task component to the next. Yet how the brain orchestrates such saccade sequences is not known. Using a multifaceted approach, we used behavioral analysis, neural recordings, computational modeling, and simulations to show that while the brain can program multiple action plans simultaneously, there are neural costs involved—the greater the overlap between the plans, the slower is the growth of neural activities encoding the two plans, leading to delayed response times. Our study sets up a neurophysiological framework to study action sequencing and is relevant for understanding sequencing errors observed in disorders like Parkinson’s disease.

Author contributions: A.M. designed research; D.B. performed research; D.B. and N.S. analyzed data; and D.B., N.S., and A.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email basu.debaleena@gmail.com.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2108922118/-DCSupplemental>.

Published October 1, 2021.

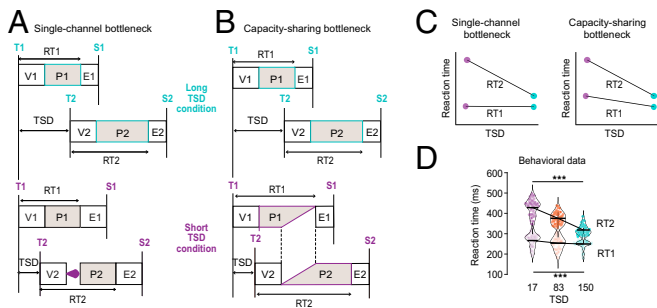


Fig. 1. Behavioral predictions for processing bottlenecks during the planning of sequential saccades. (A) Single-channel bottleneck framework. Each task is made up of three stages. The visual stage (V) can be carried on in parallel with stages of another task, but the central planning stage, P, can only proceed serially. In a two-saccade sequence, the stages of the first saccade plan proceed to completion unabated, leading to its execution (E). For the second plan, however, if the second target closely follows the first (short TSD condition), the central planning stage, P2, is postponed until P1 is complete. Such a postponement does not occur in the long TSD condition, in which the two saccade plans are well separated, thereby leading to an increase of RT2 from long to short TSD. (B) Capacity-sharing bottleneck framework. In this framework, the P stages of multiple plans can proceed in parallel and access the brain's limited processing capacity simultaneously. In the short TSD condition, P1 and P2 concurrently "share" the capacity, resulting in the slower progress of both saccade plans. This leads to the lengthening of both RT1 and RT2 in the short TSD condition, the effect on RT2 being greater as the second saccade plan gets a smaller share of the central capacity. (C) Predictions of reaction time versus TSD for single-channel bottleneck framework (Left) and capacity-sharing bottleneck framework (Right). RT2 increases with the decrease in TSD for both frameworks, whereas RT1 increase is predicted only by the capacity-sharing model. (D) Behavioral data for reaction time versus TSD. Data shows trials in which the first (for RT1) or second (for RT2) saccade was into the RF. Both reaction times increased significantly with the decrease in TSD. T1, T2, S1, S2 refer to the onsets of the first target, second target, first saccade, and second saccade, respectively. *** $P < 0.001$, ** $P < 0.01$.

Results

Two monkeys, a *Macaca radiata* (J) and a *Macaca mulatta* (G) performed a sequential saccade "FOLLOW" task (SI Appendix, Fig. S1; see SI Appendix, Methods), in which the majority (70%) of the trials were "step trials," in which they had to perform a rapid sequence of saccades to two targets in the order of their presentation. The remaining 30% of the trials were "no-step" trials, wherein a single, visual target was presented, and the monkeys had to make a single saccade to it. The two types of trials were randomly interleaved. The temporal gap (TSD) between the first and second target onsets in step trials was randomly chosen among 17, 83, and 150 ms (4).

Behavioral Evidence of Processing Bottlenecks during Sequential Saccades. In the scheme of single-channel bottleneck models, the second plan shows the hallmark of processing bottlenecks: an increase in latencies with a decrease in TSD, while saccade 1 latencies (reaction time 1 [RT1]) stay unaffected (Fig. 1C, Left). However, unlike the single-channel bottleneck model, in which plan 1 may be assumed to get 100% of the capacity, the capacity-sharing model predicts that the latencies of the first saccade (RT1) will also increase as it only gets a part of the full available capacity (Fig. 1C, Right).

To ensure that the behavioral data are matched to the neural data, we analyzed trials in which saccades were made into the response field (RF; see SI Appendix, Methods). That is, for RT1, the first saccade was made into the RF, and for RT2, the second saccade was made into the RF. Both RT1 and RT2 slowed down as the TSD decreased, indicating a capacity-sharing mechanism [Fig. 1D; RT1: Kruskal–Wallis test, $\chi^2(2, 240) = 17.85$, $P <$

0.001, and $\eta^2 = 0.07$; RT2: Kruskal–Wallis test, $\chi^2(2, 233) = 158.37$, $P < 0.001$, and $\eta^2 = 0.67$]. While the effect on RT1 was typically much smaller than that on RT2, the increases in saccade latencies with decreasing TSD corroborated with the previously well-established evidence of processing bottlenecks in concurrent action planning. Our behavioral data, thus, confirm the notion that there are bottlenecks associated with parallel programming and support the presence of a capacity-sharing bottleneck, as opposed to the single-channel bottleneck, as the first saccade plan does not stay unaffected.

Movement-Related Activity during Single Saccades. Previous work has shown that the pattern of activity of FEF movement neurons are correlated with stochastic accumulation, which is widely used in computational models of saccadic reaction times (30, 33–35) and are directly linked to saccade initiation times (36–38). Since reaction time lengthening is the main behavioral evidence of processing bottlenecks, movement neurons are well projected to carry neural correlates of the same. To confirm whether movement-related activity in FEF adheres to an accumulation-to-threshold model of reaction time, we first studied the no-step, single-saccade trials. We divided these trials into fast, medium, and slow reaction time groups and measured the parameters of accumulator models from the presaccadic activity (SI Appendix, Fig. S2A). The reaction time grouping was obtained by partitioning reaction times in each session using the mean reaction time of that session (see SI Appendix, Methods). The main parameters of accumulator models (i.e., baseline, onset, growth rate, and threshold activity) were measured for the three reaction time conditions (fast, medium, and slow) for each neuron (SI Appendix, Fig. S2 C–F; see SI Appendix, Methods). Consistent with earlier studies (30), adjustments in the growth rate of the activity of FEF movement neurons predicted reaction times in the no-step trials: Across the movement neuron population, the slope of the best fitting line for growth rate variation in the reaction time groups was significantly different from zero ($Z_{\text{rate}} = -4.27$ and $P < 0.001$; SI Appendix, Fig. S2E). Furthermore, the slopes for the growth rate were negative, indicating that fast reaction times were preceded by a steeper rate of growth of movement activity and vice versa. While the growth rate varied with reaction time, the threshold did not ($Z_{\text{threshold}} = -0.98$ and $P = 0.323$; SI Appendix, Fig. S2F), corroborating the established reaction time models of accumulation to a fixed threshold. The slope distributions of other accumulator measures, like baseline and onset, were not statistically significant from zero ($Z_{\text{baseline}} = -2.04$ and $P = 0.05$; $Z_{\text{onset}} = 1.92$ and $P = 0.054$).

Processing Bottlenecks Underlie the Representation of Sequential Saccades. Using a computational model, Sigman and Dehaene (31) had shown that evidence accumulation, representing a central decision process, constituted a bottleneck in dual tasks, while the perceptual stage and the execution stage did not. Based on the mapping between accumulator models and movement neuron dynamics, four possible hypotheses (SI Appendix, Fig. S2A) can explain how the activity of FEF movement neurons, coding for the second saccade, might bring about the systematic increase of the latency of the second saccade (RT2), with a decrease in TSD that characterizes processing bottlenecks. The lengthening of reaction time may be due to 1) the lowering of the baseline firing rate from long to short TSDs, 2) the delaying of the onset of the activity related to the second saccade from long to short TSDs, 3) the reduced growth rate of the activity with shorter TSDs, and 4) an increase of the saccade threshold firing rate from long to short TSDs.

Fig. 2 schematically shows the possible modulations of the accumulation process in the planning stage (P) and the corresponding movement neuron activity. The accumulation process is represented as a noisy integrator accumulating visual evidence until it reaches the threshold. In the single-channel bottleneck

model, RT1 is unaffected, and thus, the dynamics of the integrator and the corresponding neural activity are unchanged across the three TSDs (Fig. 2A). For RT2, the single-channel bottleneck model posits a postponement of the central stage; thus, the onset of the accumulating process and of the neural activity are delayed, as the overlap between the two saccade plans increases from long to short TSD (Fig. 2B). According to the capacity-sharing model, the first and second saccade plans can proceed in parallel; thus, there is no “waiting period” for the accumulation process of the second plan—the onset of neural activity is similar across TSDs for both first and second saccade. However, since both motor plans share the limited processing capacity, the central stages of both plans are lengthened. This could be brought about by a decrease in the rate of integration, an increase in the decision threshold, or a decrease in the baseline activity from long to short TSD. Fig. 2C and D schematically shows the hypothesized rate and threshold modulations. At the level of neural activity, the rate of ramping up movement-related activity could slow down, or the threshold firing rate for saccade onset could increase to account for the increase in saccade latencies, with a decrease in TSD. Critically, the modulations ought to be present for both saccade

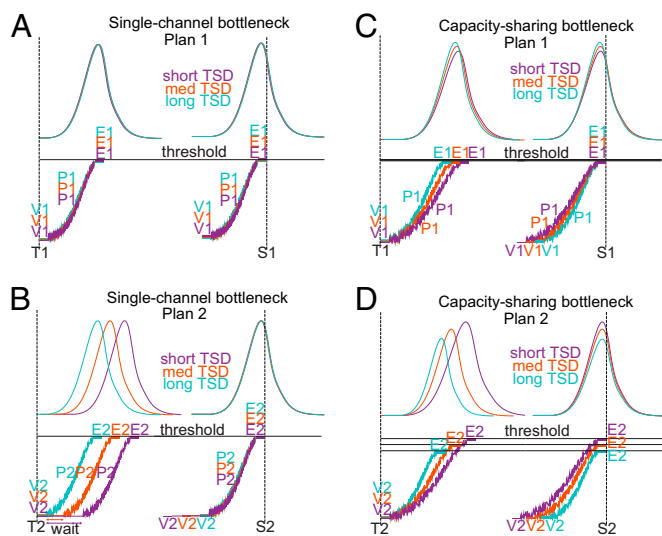


Fig. 2. Neural predictions for processing bottlenecks during the planning of sequential saccades. (A) Hypothesized neural activity for saccade plan 1 in the single-channel bottleneck framework. (Bottom) After the first visual target is presented (T1; vertical broken line), there is an initial visual processing stage (V1), which is often of constant duration for all plans. The planning stage (P1) for the first saccade is represented as a noisy integrator, accumulating activity until the motor threshold (horizontal solid line) is reached, and the saccade is executed (E1). (Top) The corresponding neural activity is shown as the ramping up of FEF movement neuron activity until saccade onset (S1). The activities corresponding to three different TSDs are shown in three different colors. (B) Hypothesized neural activity for saccade plan 2 in the single-channel bottleneck framework. The onset of the accumulation process and the ramping up of the neural activity shifts later with the decrease in TSD to account for RT2 elongation (same format as A; T2: onset of second target, V2: visual processing stage for target 2, P2: planning stage for saccade 2, E2: execution stage for plan 2, and S2: onset of second saccade). (C) Hypothesized neural activity for saccade plan 1 in the capacity-sharing bottleneck framework. The onsets of the integrators and the movement neuron activity do not change with TSD, on account of parallel programming of the two saccade plans. The increase in saccade latencies at shorter TSDs could be brought about by a decrease in the growth rate from long to short TSD, in the same format as A. (D) Hypothesized neural activity for saccade plan 2 in the capacity-sharing bottleneck framework, same as C, with the addition of threshold modulation and a greater degree of rate adjustment with TSD to constitute the larger increase in RT2 from long to short TSD, in the same format as A.

plans, according to the capacity-sharing model, although the effect would be lesser for the first plan, as the corresponding increase in RT1 is also less. While we have presented polarized scenarios for the two bottleneck models, it is possible that, at the neuronal population level, there would be a combination of the factors mentioned.

To assess which of the above possibilities explain the increase in RT2, we analyzed the neural activity in trials in which the second saccade was made into the RF for all three TSDs (Fig. 3A; see *SI Appendix*, Fig. S3 for single-neuron example). Across the population, the rate of neural activity growth slowed down from long to short TSD, and the activity ramped up to a higher firing rate threshold. We measured each of the four accumulator parameters, baseline; onset; rate; and threshold (averaged across trials of the same TSD), for the three TSD conditions for each neuron (Fig. 3B; see *SI Appendix*, *Methods*) using linear regression. The slopes from all the movement neurons were compared using a Wilcoxon signed-rank test. Across the movement neuron population, the slopes of the rate and the threshold, as a function of TSD, were significantly different from zero ($Z_{\text{rate}} = 4.27$ and $P < 0.001$; $Z_{\text{threshold}} = -2.67$ and $P < 0.01$; Fig. 3B). Furthermore, the slopes for the rate of activity growth were positive, indicating that the rate of activity grew faster at longer TSDs, in which presumably the effect of processing bottlenecks was the least among the three TSD conditions. Threshold slopes were significantly negative, indicating that, as the TSD increased, the threshold required for the initiation of the second saccade was reduced at the population level. However, the slope distributions of other accumulator measures, like baseline and onset, were not statistically significant from zero ($Z_{\text{baseline}} = -0.62$ and $P = 0.53$; $Z_{\text{onset}} = 0.86$ and $P = 0.17$). Thus, processing bottlenecks at the level of FEF movement neurons were characterized by multifaceted adjustments in the rate and threshold of the activity related to S2.

While the classical evidence of processing bottlenecks is indexed by the increase in RT2, RT1 may also be affected, according to the capacity-sharing scheme of processing bottlenecks (Fig. 1B). We tested whether movement-related activity encoding first saccade remained unchanged, as would be expected in the single-channel bottleneck scheme, or change systematically across TSDs, as the capacity-sharing model predicted. To address this issue, we performed the same analyses as before but for the trials in which the first saccade was made into the RF (Fig. 3C; see *SI Appendix*, Fig. S3 for single-neuron example). At the population level, rate perturbation occurred with a decrease in TSD in the first plan, mirroring the modulation observed for the second plan (Wilcoxon signed-rank test for slopes of rates, $Z_{\text{rate}} = 3.86$ and $P < 0.001$; Fig. 3D). However, unlike the second plan, threshold activity did not show a significantly decreasing relation with TSD ($Z_{\text{threshold}} = 1.16$ and $P = 0.25$). The slope distributions of other accumulator measures, like baseline and onset, were not statistically significant from zero ($Z_{\text{baseline}} = 0.85$ and $P = 0.39$; $Z_{\text{onset}} = 1.84$ and $P = 0.07$; Fig. 3D). Thus, rate perturbation constituted a major mechanism through which the ramping up of activity of FEF movement neurons was controlled during the parallel planning of sequential saccades.

State Space Dynamics and Inhibitory Control May Enable Capacity Sharing during Sequential Saccade Planning. To gain deeper insights into the neural mechanisms underlying capacity sharing, we studied the population dynamics underlying the trajectory of neural activity in FEF. First, we visualized this by performing a principal component analysis (PCA), separately for the population neural activity (for saccades into RF), aligned to target 1 and target 2 onsets for each of the three TSDs (Fig. 4A). PCA is a commonly used, unsupervised learning algorithm to extract latent information from the data (see *SI Appendix*, *Methods*). This method allows us to look at high-dimensional FEF population

neural activity in a much lower dimension that captures the maximum variance of the population. At least ~ 7 to 8 principal components (PCs) were required to explain $>99\%$ of the variance for any of the six conditions (three TSDs for each of the two saccade plans), although there was a trend of fewer PCs explaining more variance as TSD increased. However, the top three PCs explain $>90\%$ variance. Therefore, we visualized a “state space trajectory” by plotting the top three PCs versus one another (Fig. 4B). Each point on the trajectory indicates the neural state at each time point.

If planning for the first and the second saccades are processed in parallel but compete for the same, shared space because of limited capacity (according to the capacity sharing model), we should expect the neural trajectories to span different subspaces at shorter TSDs and span the same subspace at higher TSDs. That is, in the lowest possible TSD, we should expect the two subspaces to be completely orthogonal (no overlap), and as the TSD increases and approaches the reaction time of the first saccade, the subspaces can begin to overlap. Therefore, in our case, with the lowest TSD being 17 ms, we should expect a low degree of overlap, and at TSD = 150 ms (approximately RT1), we should expect a high degree of overlap of subspaces. In contrast, the single-channel bottleneck hypothesis predicts that the subspaces corresponding to the planning of the first and the second saccades would completely overlap, since the plan 2 would be completely dormant until plan 1 is completed.

We found that the neural trajectories significantly differed between the planning of the first and the second saccades for the shortest TSD but became more similar as the TSD increased (Fig. 4B). We quantified the degree of overlap between the subspaces spanned by these neural trajectories (see *SI Appendix, Methods*). At the shortest TSD, the magnitude of overlap between the signals for planning of the first and the second saccades was 47%, and this increased as the TSD increased from medium (84%) to long TSDs (92%; Fig. 4C), aligned more with the predictions of the capacity-sharing model. This result also held true for all saccade directions (*SI Appendix, Fig. S4*). We also confirmed that these differences were related to the TSDs and not to differences in saccade kinematics, which were similar across TSDs for the first and the second saccades (*SI Appendix, Fig. S5*).

Next, we investigated the mechanism behind the differences in the neural subspace overlap among different TSDs. We performed two sets of simulations (see *SI Appendix, Methods*; *SI Appendix, Fig. S6A–F*) using the accumulator framework (*SI Appendix, Fig. S6H*). For each of the two sets, we simulated 40 neurons with 900 trials per neuron (with three types of TSD trials) using a firing rate model to approximately match the statistical power of our experimental dataset (see *SI Appendix, Methods*). We constructed an inhibition function, such that the magnitude of the inhibition inversely varied with TSD (see *SI Appendix, Methods*; *SI Appendix, Fig. S6G*).

In the first set of simulations, we introduced a nonmutual inhibition from the saccade plan 1 to plan 2 (Fig. 5A; see *SI Appendix, Methods*). Here, the activities for plan 2 were temporally shifted by plan 1, following the inhibition curve as a function of TSD. The resulting, simulated neural activities (Fig. 5B) resembled the predictions of a single-channel bottleneck model (Fig. 2A and B). Very few (~ 3) PCs explained the $>99\%$ of the variance. The state space neural trajectories were not significantly different between the planning of the first and the second saccades for any of the TSDs (Fig. 5C), since the subspace overlap was 98% between any pair of plans (Fig. 5D), as expected from the single-channel bottleneck model.

In the next set of simulations, we introduced asymmetric mutual inhibition between the two saccade plans (see *SI Appendix, Methods*; Fig. 5E). That is, plan 1 temporally shifted plan 2 just like before, but plan 2 reduced the magnitude of peak firing of plan 1. Hence, the nature of inhibition is both mutual and asymmetric. The simulations of this model (Fig. 5F) resembled the neural data (Fig. 3A and C) and the predictions of a capacity-sharing bottleneck model (Fig. 2C and D). Here, ~ 7 to 8 PCs were required to explain $>99\%$ of the variance for the shortest TSD and fewer (~ 5 to 6) PCs were required to explain $>99\%$ of the variance for the longest TSD. The neural trajectories significantly differed for short TSD but were similar for longer TSDs (Fig. 5G), and the degree of subspace overlap between the two plans increased with TSD, consistent with the structure present in the neural data (Fig. 5H) resembling the experimental data (Fig. 4C), as expected from the capacity-sharing model.

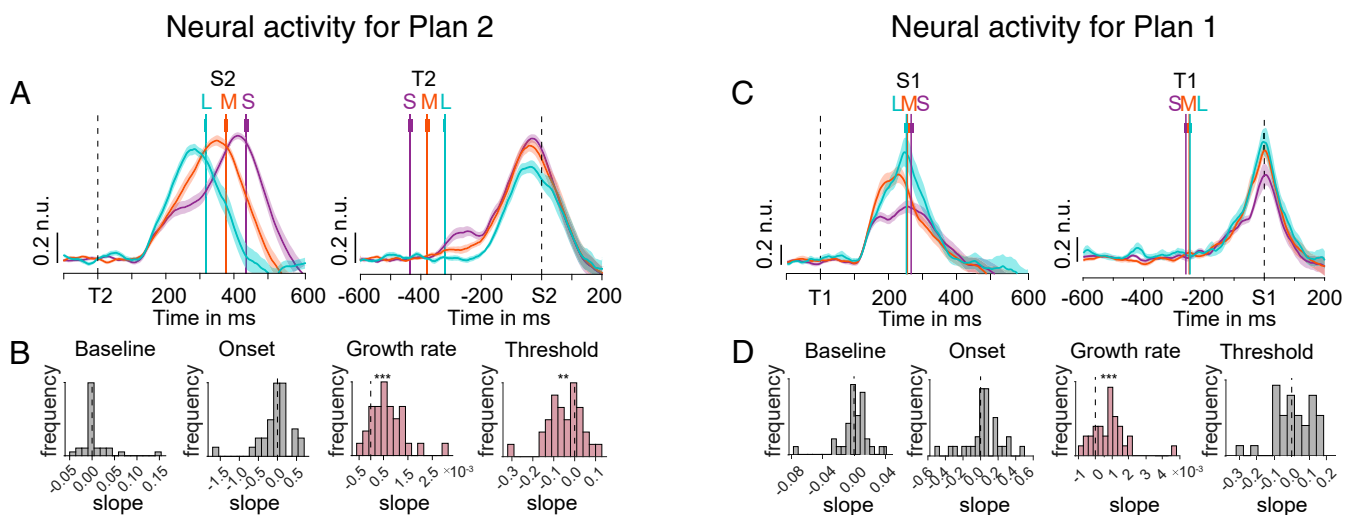


Fig. 3. Activity of FEF movement neurons during the planning of sequential saccades. (A, Left) Population activity of FEF movement neurons when the second saccade went into the RF, aligned on second target onset (T2). Mean saccade onset times (S2) for the short (S), medium (M), and long (L) TSD conditions are shown as vertical, colored lines with SEM error bars. (Right) Same as left but activity aligned to the second saccade onset. Shading indicates mean \pm SEM. n.u., normalized unit. (B) Population histogram of slopes for each measure of accumulator dynamics (baseline, onset, growth rate, and threshold) as a function of TSD for FEF movement neurons. Asterisks denote cases in which the distribution of movement neuron slopes was significantly different from zero (Wilcoxon signed-rank test, $***P < 0.001$ and $**P < 0.01$). (C) Same as A but for saccade 1. (D) Same as B but for saccade 1.

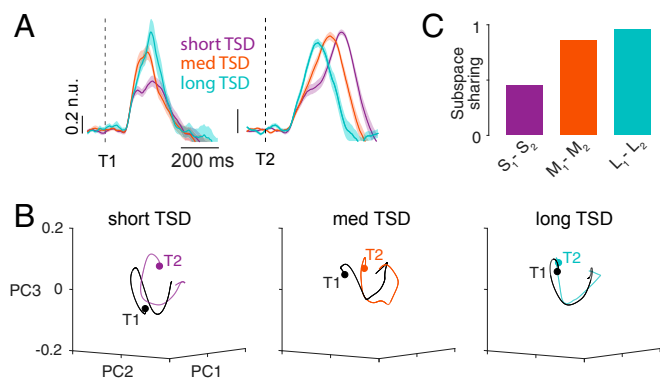


Fig. 4. Extent of subspace sharing explains processing bottlenecks during the planning of sequential saccades. (A) Normalized mean population neural responses aligned to target 1 (T1) and target 2 (T2) for short, medium, and long TSD trials (n.u. = normalized unit); this is the same as in Fig. 3 A and C. (B) The first three PCs plotted against each other for target 1 (black) and target 2 (color) related responses for short (Left), medium (Center), and long (Right) TSD trials. Filled circle markers indicate the starts of the respective trajectories. (C) Subspace overlap between a pair of conditions. S, M, and L indicate short, medium, and long TSDs, and 1 and 2 indicate the saccade plan number.

FEF Visually Related Neurons Do Not Show Processing Bottlenecks.

Previous studies have reported a separation between the visual and movement processing in FEF neurons, with only movement processing affecting reaction time in perceptually simple tasks (34, 39, 40). Thus, it is plausible that the responses of visual neurons are not gated by inhibitory bottlenecks. This notion was tested by analyzing target-related activity in purely visual (Fig. 6A) and visuomovement neurons (Fig. 6B).

We analyzed the average target-related response in the 200-ms window, following target onset for each neuron to identify the signatures of processing bottlenecks. If target selection is capacity-limited, then, presumably, neural responses encoding saccade targets appearing in close succession will be inhibited, either because of single-channel bottleneck (only second target response gets affected) or because of capacity sharing (both first and second target responses get affected). In contrast to movement-

related activity, the average activity in the target-related period did not vary with TSD for both saccade plans [Kruskal–Wallis test: $\chi^2(2, 129) = 0.47$ and $P = 0.79$ (first saccade); $\chi^2(2, 124) = 0.06$ and $P = 0.97$ (second saccade)], suggesting that the visual processing stage is prebottleneck, at least in a perceptually simple task like the FOLLOW task.

Discussion

In this study, we explored the limits of parallel processing involved in saccade sequences. For single saccades in isolation, the classical accumulator model suggests that the changes in reaction time could be brought about by changing the rate of accumulation toward a fixed activation threshold (SI Appendix, Fig. S2; 30). However, it is unclear how accumulator dynamics would underlie the reaction time variations involved in programming a sequence of saccades. We found evidence for processing bottlenecks within FEF, the mechanisms being rate perturbation and threshold modulation in the movement neuron population. Additionally, we found evidence of processing bottlenecks in the motor plans of the first and second saccades, suggesting that the associated bottleneck could be a consequence of capacity sharing between coactivated movement plans. The notion of such shared and limited processing was also revealed in the state space dynamics of FEF movement activity, which showed a potential role for inhibitory control that gated the access of concurrent motor plans to a planning subspace. Our analysis of visual activity did not reflect any consistent modulation that could be considered a significant bottleneck. The major results are discussed and interpreted in the following sections.

Processing Bottlenecks in Sequential Saccade Planning. Processing bottlenecks and parallel programming represent functionally antithetical processes, and yet both are essential for optimal saccadic behavior. While parallel programming allows for the rapid execution of a saccade sequence, processing bottlenecks are likely to arise to check unbridled parallel programming of motor plans—as failure to control it might lead to errors like averaged saccades or the incorrect order of execution of a saccade sequence (6, 32, 41–44). In the context of the current study, we tested whether a single-channel bottleneck (16) or a capacity-sharing bottleneck (22) best explained our reaction time data, as behavioral evidence

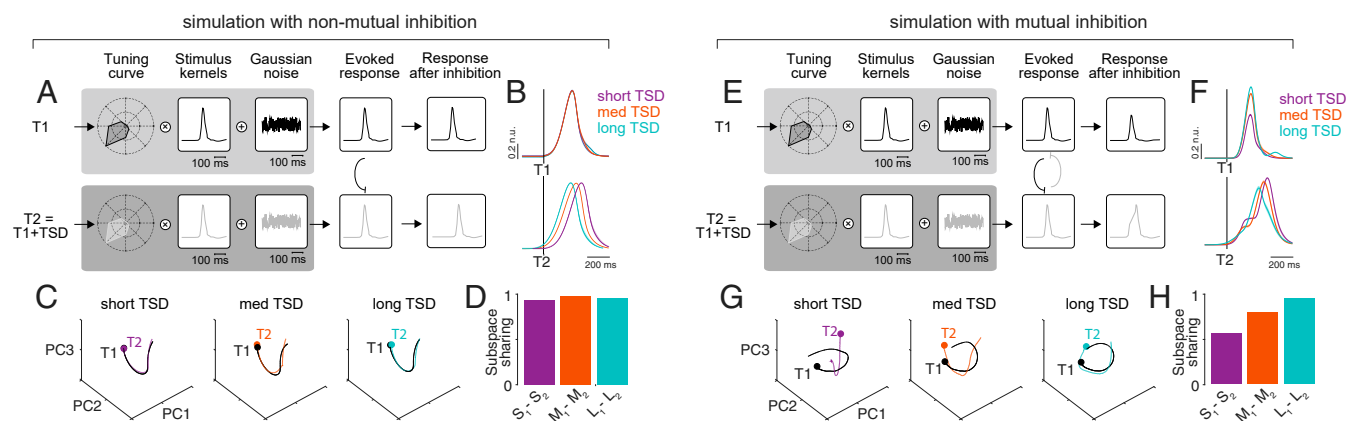


Fig. 5. Only simulations with mutual, asymmetric inhibition capture the empirical data's population dynamics. (A) Schematic of the simulation with non-mutual inhibition. (Top row) Simulated neural activity for the first saccade plan. (Bottom row) Simulated neural activity for the second saccade plan (SI Appendix, Fig. S6 and Methods; n.u. = normalized unit). (B) Normalized mean population neural responses, for data simulated with nonmutual inhibition, aligned to target 1 and target 2 for short, medium, and long TSD trials. (C) First three PCs plotted against each other for target 1 (black) and target 2 (color) related responses for short (Left), medium (Center), and long (Right) TSD trials. Filled circle markers indicate the starts of the respective trajectories. (D) Subspace overlap between a pair of conditions. S, M, and L indicate short, medium, and long TSDs, and 1 and 2 indicate the plan number. (E) Same as A but for simulation with mutual inhibition (see SI Appendix, Methods). (F) Same as B but for simulated data with mutual inhibition. (G) Same as C but for simulated data with mutual inhibition. (H) Same as D but for simulated data with mutual inhibition. T1 and T2 indicate the first and second target onsets, respectively.

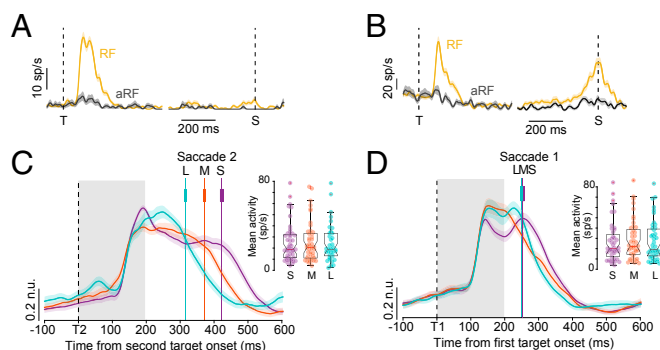


Fig. 6. Activity of FEF visual and visuomovement neurons during the planning of sequential saccades. (A) A representative FEF visual neuron aligned to target onset (T) and saccade onset (S) in a MG task for saccades into the RF (yellow) and saccades out of RF (anti-RF or aRF; black). (B) A representative FEF visuomovement neuron aligned to T and S in a MG task for saccades into the RF (yellow) and aRF (black). (C) Population activity of FEF visual and visuomovement neurons for different TSD conditions (short [S], medium [M], and long [L]) aligned to target 2 onset. Mean saccade onset times for the TSD conditions are shown as vertical, colored lines with SEM error bars. (Inset) Kruskal–Wallis box plots for average activity in the visual epoch (gray shaded area) for the three TSDs. Shading indicates mean \pm SEM. n.u., normalized unit. (D) Same as C but for neural activity aligned to target 1.

of both the models have been found in dual-task paradigms (16, 25, 45). In our data, we found evidence of an increase in both RT1 and RT2 with TSD, ruling out the possibility of the single-channel bottleneck model being the exclusive framework underlying bottlenecks in sequential saccades. Our neural data also suggested a capacity-sharing mechanism of bottlenecks: The onset of saccade-related activity did not vary with TSD, negating the single-channel bottleneck hypothesis (see Fig. 2), and both saccade plans showed consistent activity modulations with TSD. A reduction in the rate of accumulation and an increase in the threshold activity level were seen for the second saccade plan. In contrast, only changes in the slope of the activity corresponding to the first saccade were observed, which may account for the subtler changes in RT1.

A question that arises is whether the bottleneck between the first saccade and the second saccade can have a component of an aborted, retinotopic saccade to the final target from the fixation point. In the FOLLOW task, the first saccade is made to the retinotopic location of the first target, whereas the second saccade is toward the remapped location of the second target, with respect to the first. We analyzed only correct trials in this study, in which accurate second saccade mapping had obviously occurred. Any covert saccade plan to the retinotopic location of the second target would have to be quickly cancelled, before the monkey made the first saccade to the initial target while planning remapped second saccade. Our previous study indicated that the remapped, and not the retinotopic second saccade vector is concurrently planned with the first saccade (4). This observation is corroborated in the current study by the pattern of reaction times and changes in accumulation parameters that suggest that it is the remapped saccade, which is affected by the concurrent preparation of the first saccade. However, it is plausible that a covert and aborted saccade plan from the fixation point to the second target is also a source of interference caused by the parallel programming of the two saccade plans, which are ultimately executed.

While investigating the limits of parallel programming of saccade sequences, it is important to note that the longer latencies of the second saccade plan do not preclude concurrent saccade programming. Our previous study has shown evidence of parallel programming of saccades in the same dataset (4). The longer

saccade latencies of second saccades in the FOLLOW task has been consistently observed in previous studies (3, 32). Previous studies have established that the degree of parallel programming is best measured by the extent of the intersaccadic intervals (ISIs; 1, 3, 5, 9). For purely serial programming, the gap between a pair of saccades or the ISI would be equivalent to the single saccade RT, since planning for the second saccade will start after the first is completed. In case of parallel programming, the ISI would be lesser, indicating that the processing of the consecutive plan begins while the current plan is underway. In our dataset, the first saccade RTs approach no-step RTs across all TSDs (SI Appendix, Fig. S7A). In the serial programming condition, the total sequence time would be double that of the first saccade reaction time. The total sequence time obtained experimentally would span the length of the first saccade reaction time and the ISI. SI Appendix, Fig. S7B shows that the serialized sequence time is longer than the observed total sequence time [ANOVA, $F(1, 48,882) = 20,344.24$ and $P < 0.001$], indicating parallel programming. As shown in Fig. 1D, there are limits to parallel programming, especially for closely overlapping saccade plans (i.e., TSD 17 ms), wherein processing bottlenecks lengthen both the RTs, with the effect being more pronounced for the second saccade as compared to the first. In addition, the longer second saccade RTs may also arise from a consideration of the ergodic nature of saccade planning, in which saccade variability is dominated by the variability across trials (46). Here, once a rate is set on a given trial, it determines the planning trajectory and the ensuing RT for that trial. Seen in this context, even if two saccade plans are proceeding in parallel, inhibition that is capable of decreasing the rate of accumulation can lead to longer second saccade RTs, because of slower growth in activity that persists even after the first saccade is executed.

Inhibitory Control Underlying Processing Bottlenecks. We tested whether mutually inhibitory accumulators encoding distinct saccade plans can mimic capacity sharing, wherein both the saccadic eye movements are executed with delays, especially for the second saccade. Modeling such a response required two important conditions: The first condition required that the inhibition be asymmetric, being greater for the first saccade plan than the second saccade plan, which manifest as greater capacity and faster information processing for the former, compared to the latter. Such an asymmetry is a natural consequence of the temporal delay, allowing for greater activity in the first saccade to inhibit the second saccade; the second condition required an inhibitory kernel that decreased with TSD, such that inhibition from the first accumulator is greater at shorter delays, despite the level of activity in the accumulator being lesser compared to what it would be at larger TSDs. Such an inhibitory kernel is necessary to match the observed behavioral data of greater second saccade reaction times as well as the neural data, which showed greater interference for the second saccade motor plans at the shorter TSDs. Interestingly, using a dynamical systems approach under the assumption of stationarity of noise across trials (47), this model of inhibitory control could be also shown to act as a “queuing” mechanism, in which nonorthogonal, neural spaces can simultaneously allow parallel processing but yet temporarily slow the processing of the second saccade. We believe that the ability of such inhibition to reconfigure the neural space may reflect the nonlinear effects of inhibition on the pattern of activity, representing accumulator activity that underlie the saccade planning.

The simplest and most parsimonious explanation for the location of such a bottleneck is at the level of FEF via mutual inhibition (48) of competing motor plans developing in the FEF. This type of inhibitory gating can be brought about by inhibitory interneurons within the FEF (49, 50). Although such a form of inhibition is intuitive and can be readily implemented within the proposed frameworks described for decision-making circuits (51, 52), implementing an inhibitory kernel that decreases with

increasing TSD cannot be easily implemented in a straightforward manner by mutually inhibitory accumulators. Furthermore, using an identical task, our previous work has shown that the basal ganglia (BG) is causally involved in the conversion of parallel movement plans into sequential behavior (7). The inactivation of the BG in monkeys with muscimol or impairment of the BG in Parkinson's disease patients resulted in a significantly greater extent of saccadic errors that develop because of unchecked parallel programming, leading to a "collision" of saccade plans. The results of both these studies can be reconciled by the fact that FEF and BG share a closed connection through the cortico-BG-thalamo-cortical loop, wherein the thalamus, a major relay center, receives projections from BG output nuclei and, in turn, projects to multiple cortical regions, including the FEF, which are again routed to the input nuclei of BG (53–56). Thus, the origin of the bottleneck could also be in the well-established, inhibitory control circuitry of the BG (57) and then rerouted to the FEF through the BG-thalamo-cortical loop (58), which then manifests in the various adjustments of movement-related neuronal activity.

Neural Representations of Processing Bottlenecks within FEF. Our data show robust signatures of processing bottlenecks involving rate and threshold adjustments of FEF movement neurons contributing to the observed processing bottlenecks. Interestingly, similar adjustments of rate have been observed in FEF movement-related neurons, when monkeys slow their reaction times to improve their accuracy (59), consistent with movement-related activity reflecting a developing motor plan that can be adjusted by strategic requirements of the task. However, in contrast to speed/accuracy adjustments, we did find systematic increases in threshold for the second saccade with shorter TSDs that, together with decreases in accumulation rate, contribute to the lengthening of reaction times for the second saccade. Similar changes in growth rate for both the first and second saccade, particularly at shorter TSDs, were also observed in our model of mutually inhibiting accumulators but without any changes in threshold (Fig. 5), raising the possibility that these changes may involve additional processes, such as adjustments in the excitability of superior colliculus neurons from the BG (60, 61), that were not modeled here.

In contrast to the movement neurons, the activity of visual neurons displayed little evidence of active inhibitory control,

suggesting that they are "prebottleneck." This is not surprising since many studies have reported a separation between the visual and motor processing of FEF neurons, with only motor processing affecting reaction time in perceptually simple tasks; thus, it is plausible that the responses of visual neurons are not gated by inhibitory bottlenecks for our task. However, it can be speculated that, in a more perceptually challenging task, manifestations of processing bottlenecks would show up in the activity of visual responses as well. Thus, it is likely that movement neurons, which are thought to be functionally downstream of visual neurons (34), are subjected to a greater degree of inhibitory control, possibly due to its direct role in saccade initiation. A similar result was also observed in the countermanding (62) and redirect tasks (63), in which movement-related neurons showed the strongest evidence of inhibitory control that reflected the monkeys' abilities to withhold or change saccade plans. Thus, movement-related activity would fall under the post/peribottleneck category while visually related activity would be prebottleneck, at least for perceptually simple tasks.

Methods

The detailed methods pertaining to this dataset has been published in a previous study (4, 27). A brief overview is given in the *SI Appendix*. Single-unit recordings were done from two adult monkeys (J, male *M. radiata*, and G, female *M. mulatta*). The animals were cared for in accordance with the animal ethics guidelines of the Committee for the Purpose of Control and Supervision of Experiments on Animals, the Government of India, and the Institutional Animal Ethics Committee of the Indian Institute of Science that approved the protocols. Data analysis was done using custom-written codes in MATLAB (MathWorks).

Data Availability. Physiological data reported in this study have been deposited in the Mendeley Data repository (<https://doi.org/10.17632/c2rg5f92yg.1>). All data related to the paper are included in the article and/or in the *SI Appendix*.

ACKNOWLEDGMENTS. We thank S. Sengupta for helping with behavioral training and A. Gopal P.A. for helping with data collection. This work was supported by a Department of Biotechnology, Government of India—Indian Institute of Science partnership grant given to A.M. D.B. was supported by a graduate fellowship from the Ministry of Human Resource Development, Government of India, through the Indian Institute of Science. N.S. was partially supported by NIH grant R01NS 113078-01, which was given to Michael E. Goldberg.

- W. Becker, R. Jürgens, An analysis of the saccadic system by means of double step stimuli. *Vision Res.* **19**, 967–983 (1979).
- A. W. Minken, A. J. Van Opstal, J. A. Van Gisbergen, Three-dimensional analysis of strongly curved saccades elicited by double-step stimuli. *Exp. Brain Res.* **93**, 521–533 (1993).
- S. Ray, J. D. Schall, A. Murthy, Programming of double-step saccade sequences: Modulation by cognitive control. *Vision Res.* **44**, 2707–2718 (2004).
- D. Basu, A. Murthy, Parallel programming of saccades in the macaque frontal eye field: Are sequential motor plans coactivated? *J. Neurophysiol.* **123**, 107–119 (2020).
- K. M. Sharika, A. Ramakrishnan, A. Murthy, Control of predictive error correction during a saccadic double-step task. *J. Neurophysiol.* **100**, 2757–2770 (2008).
- N. Bhutani, S. Ray, A. Murthy, Is saccade averaging determined by visual processing or movement planning? *J. Neurophysiol.* **108**, 3161–3171 (2012).
- N. Bhutani et al., Queuing of concurrent movement plans by basal ganglia. *J. Neurosci.* **33**, 9985–9997 (2013).
- R. M. McPeck, J. H. Han, E. L. Keller, Competition between saccade goals in the superior colliculus produces saccade curvature. *J. Neurophysiol.* **89**, 2577–2590 (2003).
- R. M. McPeck, A. A. Skavenski, K. Nakayama, Concurrent processing of saccades in visual search. *Vision Res.* **40**, 2499–2516 (2000).
- R. M. McPeck, E. L. Keller, Superior colliculus activity related to concurrent processing of saccade goals in a visual search task. *J. Neurophysiol.* **87**, 1805–1815 (2002).
- E. X. W. Wu et al., Parallel programming of saccades during natural scene viewing: Evidence from eye movement positions. *J. Vis.* **13**, 17 (2013).
- J. Tian, J. Schlag, M. Schlag-Rey, Testing quasi-visual neurons in the monkey's frontal eye field with the triple-step paradigm. *Exp. Brain Res.* **130**, 433–440 (2000).
- A. N. Phillips, M. A. Segraves, Predictive activity in macaque frontal eye field neurons during natural scene searching. *J. Neurophysiol.* **103**, 1238–1252 (2010).
- N. L. Port, R. H. Wurtz, Sequential activity of simultaneously recorded neurons in the superior colliculus during curved saccades. *J. Neurophysiol.* **90**, 1887–1903 (2003).
- K. Shen, M. Paré, Predictive saccade target selection in superior colliculus during visual search. *J. Neurosci.* **34**, 5640–5648 (2014).
- H. Pashler, Dual-task interference in simple tasks: Data and theory. *Psychol. Bull.* **116**, 220–244 (1994).
- R. Marois, J. Ivanoff, Capacity limits of information processing in the brain. *Trends Cogn. Sci.* **9**, 296–305 (2005). Correction in: *Trends Cogn. Sci.* **9**, 415 (2005).
- A. T. Welford, The 'psychological refractory period' and the timing of high-speed performance—a review and a theory. *Br. J. Psychol.* **43**, 2–19 (1952).
- E. Ruthruff, H. E. Pashler, A. Klaassen, Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychon. Bull. Rev.* **8**, 73–80 (2001).
- A. T. Welford, Single-channel operation in the brain. *Acta Psychol. (Amst.)* **27**, 5–22 (1967).
- D. E. Broadbent, *Decision and Stress* (Academic Press, 1971).
- D. Kahneman, *Attention and Effort* (Citeseer, 1973).
- P. McLeod, A dual task response modality effect: Support for multiprocessor models of attention. *Q. J. Exp. Psychol.* **29**, 651–667 (1977).
- D. Gopher, D. Navon, How is performance limited: Testing the notion of central capacity. *Acta Psychol.* **46**, 161–180 (1980).
- D. Navon, J. Miller, Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognit. Psychol.* **44**, 193–251 (2002).
- M. Tombu, P. Jolicoeur, A central capacity sharing model of dual-task performance. *J. Exp. Psychol. Hum. Percept. Perform.* **29**, 3–18 (2003).
- N. Sendhilnathan, D. Basu, M. E. Goldberg, J. D. Schall, A. Murthy, Neural correlates of goal-directed and non-goal-directed movements. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2006372118 (2021).
- N. Sendhilnathan, D. Basu, A. Murthy, Assessing within-trial and across-trial neural variability in macaque frontal eye fields and their relation to behaviour. *Eur. J. Neurosci.* **52**, 4267–4282 (2020).
- N. Sendhilnathan, D. Basu, A. Murthy, Simultaneous analysis of the LFP and spiking activity reveals essential components of a visuomotor transformation in the frontal eye field. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6370–6375 (2017).

30. D. P. Hanes, J. D. Schall, Neural control of voluntary movement initiation. *Science* **274**, 427–430 (1996).
31. M. Sigman, S. Dehaene, Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biol.* **3**, e37 (2005).
32. S. Ray, N. Bhutani, A. Murthy, Mutual inhibition and capacity sharing during parallel preparation of serial eye movements. *J. Vis.* **12**, 17 (2012).
33. R. Ratcliff, Y. T. Hasegawa, R. P. Hasegawa, P. L. Smith, M. A. Segraves, Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *J. Neurophysiol.* **97**, 1756–1774 (2007).
34. G. F. Woodman, M.-S. Kang, K. Thompson, J. D. Schall, The effect of visual search efficiency on response preparation: Neurophysiological evidence for discrete flow. *Psychol. Sci.* **19**, 128–136 (2008).
35. L. Boucher, T. J. Palmeri, G. D. Logan, J. D. Schall, Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychol. Rev.* **114**, 376–397 (2007).
36. M. F. Huerta, L. A. Krubitzer, J. H. Kaas, Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys: I. Subcortical connections. *J. Comp. Neurol.* **253**, 415–439 (1986).
37. T. P. Langer, C. R. Kaneko, Brainstem afferents to the oculomotor omnipause neurons in monkey. *J. Comp. Neurol.* **295**, 413–427 (1990).
38. M. A. Segraves, Activity of monkey frontal eye field neurons projecting to oculomotor regions of the pons. *J. Neurophysiol.* **68**, 1967–1985 (1992).
39. K. G. Thompson, N. P. Bichot, J. D. Schall, Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *J. Neurophysiol.* **77**, 1046–1050 (1997).
40. T. Sato, A. Murthy, K. G. Thompson, J. D. Schall, Search efficiency but not response interference affects visual selection in frontal eye field. *Neuron* **30**, 583–591 (2001).
41. J. M. Findlay, Global visual processing for saccadic eye movements. *Vision Res.* **22**, 1033–1045 (1982).
42. P. Viviani, R. G. Swenson, Saccadic eye movements to peripherally discriminated visual targets. *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 113–126 (1982).
43. C. Coëffé, J. K. O'Regan, Reducing the influence of non-target stimuli on saccade accuracy: Predictability and latency effects. *Vision Res.* **27**, 227–240 (1987).
44. D. Zambambieri, R. Schmid, J. Ventre, "Saccadic eye movements to predictable visual and auditory targets" in *Eye Movements from Physiology to Cognition*, J. K. O'Regan, A. Levy-Schoen, Eds. (Elsevier, 1987), pp. 131–140.
45. K. M. Arnell, J. Duncan, Separate and shared sources of dual-task cost in stimulus identification and response selection. *Cognit. Psychol.* **44**, 105–147 (2002).
46. B. A. J. Reddi, R. H. S. Carpenter, The influence of urgency on decision time. *Nat. Neurosci.* **3**, 827–830 (2000).
47. G. F. Elsayed, A. H. Lara, M. T. Kaufman, M. M. Churchland, J. P. Cunningham, Reorganization between preparatory and movement population responses in motor cortex. *Nat. Commun.* **7**, 13239 (2016).
48. S. Ray, P. Pouget, J. D. Schall, Functional distinction between visuomovement and movement neurons in macaque frontal eye field during saccade countermanding. *J. Neurophysiol.* **102**, 3091–3100 (2009).
49. P. Somogyi, A specific 'axo-axonal' interneuron in the visual cortex of the rat. *Brain Res.* **136**, 345–350 (1977).
50. H. Markram et al., Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* **5**, 793–807 (2004).
51. R. Ratcliff, P. L. Smith, A comparison of sequential sampling models for two-choice reaction time. *Psychol. Rev.* **111**, 333–367 (2004).
52. R. Bogacz, E. Brown, J. Moehlis, P. Holmes, J. D. Cohen, The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
53. G. E. Alexander, M. R. DeLong, P. L. Strick, Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* **9**, 357–381 (1986).
54. A. Parent, L.-N. Hazrati, Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res. Brain Res. Rev.* **20**, 91–127 (1995).
55. A. Parent, L.-N. Hazrati, Functional anatomy of the basal ganglia. II. The place of subthalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Res. Brain Res. Rev.* **20**, 128–154 (1995).
56. F. A. Middleton, P. L. Strick, Basal ganglia and cerebellar loops: Motor and cognitive circuits. *Brain Res. Brain Res. Rev.* **31**, 236–250 (2000).
57. O. Hikosaka, Y. Takikawa, R. Kawagoe, Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.* **80**, 953–978 (2000).
58. P. S. Goldman-Rakic, L. J. Porrino, The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. *J. Comp. Neurol.* **242**, 535–560 (1985).
59. R. P. Heitz, J. D. Schall, Neural mechanisms of speed-accuracy tradeoff. *Neuron* **76**, 616–628 (2012).
60. C.-C. Lo, X.-J. Wang, Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat. Neurosci.* **9**, 956–963 (2006).
61. R. H. Wurtz, O. Hikosaka, Role of the basal ganglia in the initiation of saccadic eye movements. *Prog. Brain Res.* **64**, 175–190 (1986).
62. D. P. Hanes, W. F. Patterson II, J. D. Schall, Role of frontal eye fields in countermanding saccades: Visual, movement, and fixation activity. *J. Neurophysiol.* **79**, 817–834 (1998).
63. A. Murthy, S. Ray, S. M. Shorter, J. D. Schall, K. G. Thompson, Neural control of visual search by frontal eye field: Effects of unexpected target displacement on visual selection and saccade preparation. *J. Neurophysiol.* **101**, 2485–2506 (2009).