HHS Public Access

Author manuscript

Nat Neurosci. Author manuscript; available in PMC 2019 May 26.

Published in final edited form as:

Nat Neurosci. 2018 December ; 21(12): 1651–1655. doi:10.1038/s41593-018-0271-5.

Midbrain activity can explain perceptual decisions during an attention task

James P. Herman[*], **Leor N. Katz**, and **Richard J. Krauzlis**

Laboratory of Sensorimotor Research, National Eye Institute, NIH, Bethesda, MD, USA

Abstract

We introduce a decision model that interprets the relative levels of moment-by-moment spiking activity from the right and left superior colliculus to distinguish relevant from irrelevant stimulus events. The model explains detection performance in a covert attention task, both in intact animals and when performance is perturbed by causal manipulations. This provides a specific example of how midbrain activity could support perceptual judgments during attention tasks.

Decision-making is often described by models in which an abstract decision variable crosses a boundary[1]. Interpreting neuronal activity itself as a decision variable has established a compelling link between boundary-crossing models and perceptual choice behavior[2]. However, efforts to test this link by causally perturbing neuronal activity and explaining the changes in choice behavior have been largely unsuccessful; for example, causal manipulation of activity in area LIP can affect reaction time but does not change perceptual choices[3,4].

A good candidate for testing the link between boundary-crossing models and behavior is the primate superior colliculus (SC), a retinotopically organized midbrain structure. Neurons in the SC have activity related to target probability[5] and comprise a “priority” or “saliency” map[6] of the visual field. In addition, we recently found that the activity of single SC neurons is correlated with behavior in a covert color-change detection task[7]. Furthermore, perturbation of SC activity reliably alters perceptual choices in attention tasks: microstimulation causes a spatially specific increase in hits[8,9], and inactivation causes both a decrease in hits inside the affected portion of visual field and an increase in false alarms outside it[10,11]. Thus, SC neuronal activity – intact or perturbed – might be expected to predict the outcome of decisions about whether a relevant (cued) or irrelevant (un-cued) stimulus event has occurred. To test this hypothesis, we implemented a model with a decision variable based on SC neuronal activity and tested whether it could account for the

[*]Correspondence should be addressed to James P. Herman (hermanj@gmail.com), or Richard J. Krauzlis (richard.krauzlis@nih.gov).

Author contributions

J.P.H. and R.J.K. designed the experiments and model. J.P.H. and L.N.K. conducted the experiments, J.P.H. implemented the model. J.P.H., L.N.K, and R.J.K. wrote the manuscript.

Competing interests

The authors declare no competing interests.

pattern of altered choices observed during microstimulation or inactivation in a covert attention task.

Neuronal and behavioral data were collected while monkeys performed a covert color-change detection task[7] (figure 1a). The task required monkeys to release a joystick in response to subtle saturation changes at a relevant (cued) location and ignore changes at an irrelevant (un-cued) foil location. The detection was performed covertly – monkeys were not allowed to look directly at the peripheral stimuli and instead maintained central fixation throughout. In constructing the decision variable, we included data only from visual-movement and visual-movement prelude units (114 in monkey 1, 37 in monkey 2), having previously found that the activity of these unit-types best predicts choices in our task[7].

The fundamental assumption in our model is that thresholding the difference in activity of pooled left and right SC neurons is sufficient to account for the choice to either respond (by releasing a joystick) or not in the covert attention task (figure 1b). Pooled activity was the sum of single-trial activity from individual neurons (figure 1c). We visualized SC output as a "trajectory" in 2D space by plotting the pooled left SC activity against the pooled right SC activity (figure 1d). In this format, the trajectory lingers along the identity line when the difference in activity between left and right SC is small (as is evident for stimulus-onset-evoked activity) but lunges out along one of the two axes when the difference is large, reaching a maximum value at one specific time-point during the phasic activity evoked by the stimulus change (compare time-point "d" on the figure 1c and 1d). The values of these maximum excursion "summary points" of single-trial SC activity are the primary inputs to the model.

The summary points can be represented interchangeably in 2D activity space or a 1D difference-of-activity space (figure 1e, see also figure S1). We consider each. In 2D space, the model's choice to either respond or not in each trial was determined by whether the trajectory to the summary point crossed a unity-slope decision boundary (dashed diagonals in figure 1e). A pair of decision boundaries positioned symmetrically above and below the identity line allowed the model's response to be triggered by high relative activity in either the right or left SC. In 1D space, the 2D summary point '(x, y)' is equivalent to a 1D difference value 'x - y', and the decision depends on whether this difference exceeds a threshold (equivalent to the intercept of the 2D boundary). In this way, model response rates can be seen as either the percentage of summary points lying outside the pair of 2D decision boundaries or the percentage of summary point differences exceeding a 1D criterion. If the absolute value |x - y| is used in the 1D case, response rate calculations are reduced to a familiar signal detection theory style of presentation[12] (figure 1e, right). While the 1D format is elegant and compact, the 2D "activity space" format offers an advantage over the 1D space: it graphically illustrates how the summary points are affected by correlated variability between and within pools of SC neurons, which we measured and incorporated into our model following established methods[13].

We first tested the ability of the model to reproduce attention task performance in control conditions when neuronal activity was not perturbed. One decision boundary was fit to a pair of hit and false alarm rates from each recording session (n = 59 in monkey 1, n = 70 in

monkey 2). The fitted model reproduced each monkey's overall hit and false alarm rates nicely, yielding rates that are comparable in both mean and variance (figure 1f). To quantitatively compare data to model we used a generalized linear model (GLM). Examining the significance of the GLM's primary and interaction coefficients showed: (i) that monkey and model performance were indistinguishable ($p = 0.99$), (ii) that our model had done no better or worse at reproducing hits or false alarms ($p = 0.89$), (iii) nor had it done better or worse at predicting monkey 1 or monkey 2's performance ($p = 0.96$).

We next examined the factors that contributed to the model's success. Both task performance (hit and false alarm rates) and the SC neuronal activity used to model performance can be characterized by signal detection theory's sensitivity index[12]: d'. The comparison of SC activity-d' and behavioral-d' illustrates how the model is constrained by the data: it cannot produce any arbitrary pair of hit rate (H) and false alarm rate (F) values, but only those that satisfy the equation activity-d' = Z(H) – Z(F)[14] (figure 1g).

What determines the activity-d' arising from SC neuronal data? For a pair of distributions, d' is the difference in their means ($\Delta\mu$) divided by their average variance ($\bar{\sigma}$; figure 1e, right). $\Delta\mu$ depended simply on the averaged activity during cue and foil changes. $\bar{\sigma}$, however, was affected by three factors: (1) the average magnitude of trial-to-trial correlations within an SC pool, (2) correlations between left and right SC pools, and (3) the size of the SC pools. Since $\Delta\mu$ was fixed by the data, d' of SC activity depended entirely on these three factors that determine $\bar{\sigma}$, which we now consider in turn.

Correlations within (ρ_w) and between (ρ_b) right and left SC pools have opposing effects on activity-d', and the effect of adding neurons is limited to smaller pools (figure 2a). For uncorrelated neurons d' increases without bound as a function of pool size, but once correlations are introduced activity-d' asymptotes[13,15,16] at approximately 300–1000 units (figure 2a). Because of this asymptotic behavior, the pool size we chose (n = 1000 each) will yield the same results as any other size in the asymptotic regime. For a fixed pool size, increasing ρ_w decreases activity-d' by increasing $\bar{\sigma}$ (figure 2b, top). Unlike ρ_w, which is strongly skewed to positive values, ρ_b can be equally positive or negative. When $\rho_b > 0$ (right and left SC are positively correlated) variability in the difference between left and right SC decreases because shared variation is subtracted away, decreasing $\bar{\sigma}$ and increasing d'. When $\rho_b < 0$, the opposite happens, $\bar{\sigma}$ increases and d' decreases (figure 2b, bottom). Thus even for a fixed pool size, there are an arbitrary number of possible combinations of ρ_w and ρ_b that will yield the same d'.

To illustrate this point, we systematically quantified how various combinations of neuronal correlations would be expected to influence the match between activity-d' and behavioral-d'. Separately for each of the two monkeys, we considered an empirically plausible range of ρ_w values[17] (0.0 – 0.2), all possible ρ_b values for each ρ_w (given the type of covariance we assumed), and fixed the pool size at n = 1000 (each). For each combination of ρ_w and ρ_b values, we computed the activity-d' and subtracted the monkey's behavioral-d', plotting a heatmap of that difference (figure 2c). A "trough" of ρ_w and ρ_b combinations that yield the minimum difference between activity-d' and behavioral-d' are optimal for the model to predict monkey behavior.

How do the model-optimal correlation values compare to those of actual neuronal data? To measure correlations within and between pools of SC neurons, we recorded from ensembles of neurons across the right and left SCs simultaneously with a pair of linear multi-contact probes (figure 2d). Consistent with our predictions (figure 2c), we found small positive average correlations within and between pools of SC neurons (figure 2e). The average correlation values were $\rho_w = 0.09$ within pools of SC neurons and $\rho_b = 0.047$ between pools; both were significantly greater than 0 (t-tests, both $p << 0.01$) and fell within the heatmap trough of optimal values (figure 2c). The bilateral recording sessions also afforded us the opportunity to estimate the relationship between boundary-crossing time in the model and the monkey's joystick-release time (figure S3); the strong correlation we found ($r = 0.74$) lends additional support to the model.

If the model describes the SC's causal contribution to task performance, it should also account for performance changes resulting from experimental manipulation of SC activity. To test this proposal, we unilaterally manipulated SC activity (in separate sessions) with either pharmacological inactivation or electrical microstimulation, following protocols known to produce behavioral effects[8,11] (figure 3a). Inactivation caused decreases in hit rate inside the affected area of visual space and increases in false alarm rate outside it, reducing behavioral-d' (mean reduction = −1.6, 95% CI = [−1.9, −1.2]) and increasing criterion (0.17, [0.1, 0.3]); microstimulation had the opposite effects (d': 0.3, [0.2, 0.5]; criterion: −0.3, [−0.4, −0.2]).

We simulated the effects of manipulating SC activity by adding one additional parameter that multiplicatively scaled up or down the output activity of one SC (figure 3b).

Multiplicative scaling with no change of decision boundary allowed the model to successfully reproduce the behavioral effects of inactivation and microstimulation in both monkeys. During simulated inactivation, scaling down right SC activity compressed summary points horizontally in 2D activity space, moving 1D difference distributions closer (figure 3b); the resulting reduction in activity-d' predicts a decreased hit rate and an increased false alarm rate, even without changing the decision boundary (figure 3c, left). Conversely during microstimulation, scaling up right SC activity horizontally spread out summary points, moving 1D distributions further apart, increasing activity-d' and thereby predicting improvements in performance (figure 3c, right).

As alternative explanations to multiplicative scaling of the SC activity with a constant boundary, we also considered additive scaling (with a constant boundary), as well as multiplicative scaling with a variable boundary. Performance of the additive scaling model was significantly worse than multiplicative scaling models at simulating inactivation but equivalent for microstimulation, and a variable boundary offered no improvement over a constant boundary with multiplicative scaling (post-hoc testing with $\alpha = 0.05$ following an ANOVA). This result indicates that multiplicative scaling of SC activity in our pooling model was sufficient to account for the effects of causal manipulations of SC activity (see also figure S4). Finally, we subjected the model to an additional test: we exploited idiosyncratic differences in behavior exhibited by the two monkeys during activity-perturbation sessions to test whether model success was specifically tied to each monkey's

SC activity. We repeated the process of model fitting but used the activity from monkey 1 to predict performance of monkey 2, and vice-versa. This “activity swap” test produced error distributions with significantly larger means (but similar variances) than in the un-swapped cases (bootstrap tests, all $p < 0.01$, variances: $p = 0.05–0.09$), showing that good fits provided by the model were not guaranteed or arbitrary, but instead were dependent on the particular SC activity recorded in each monkey.

In summary, our results illustrate the predictive relationship between SC output and performance in a covert detection task – both in intact animals (using a 1-parameter model) and during sessions where performance was altered by experimental manipulation of SC activity (2-parameters). The SC pooling model also explains why unilateral manipulations cause bilateral behavioral effects: because relevant events are detected based on the relative levels of activity in the right and left SC, unilateral manipulation of activity affects performance for events in either visual field. We acknowledge that there are other possible models and ways to simulate changes in SC activity that we have not considered. Our results are agnostic about detailed mechanisms but drive intriguing hypotheses. For example, the comparison of the relative levels of SC activity could be implemented as a normalization-like computation[18], allowing it to operate simultaneously over stimulus events across the entire visual field rather than requiring a mechanism that explicitly compares pairs of visual field locations.

Together, this work demonstrates a biologically plausible mechanism by which changes in subcortical signals for perceptual decision-making can give rise to some of the behavioral correlates of spatial attention. Our model shows how cue-related modulation of SC activity could influence perceptual choices by affecting boundary-crossing probability (figure 1e). For detection tasks, the signals read out from the SC alone can be sufficient to explain behavioral responses, as demonstrated. For discrimination tasks, because primate SC neurons are, at most, weakly tuned for visual features, boundary-crossing would require incorporating sensory information from cortex and elsewhere to correctly guide perceptual decisions. However, altering the probability of boundary-crossing in SC would alter response probability independent of the quality of additionally available sensory information; this explains how inactivation of SC can eliminate the effects of spatial cues on perceptual sensitivity[19] even though SC activity does not represent the discriminated feature or influence the fidelity of local visual signals[11,20]. Our results Illustrate how decision-making and selective attention are intertwined and highlight the importance of understanding how signals from the SC, as well as the cortex, are read out and interact during perceptual choices.

Online Methods

General

Data were collected and analyzed from two adult male rhesus monkeys (*Macaca mulatta*) weighing 9–12 kg, and were not performed blind to the conditions of the experiments. All experimental protocols were approved by the National Eye Institute Animal Care and Use Committee and all procedures were performed in accordance with the United States Public Health Service policy on the humane care and use of laboratory animals.

Task

The details of our covert color-change detection task and dynamic color stimuli have been described in detail previously[7] (figure 1c). Briefly, each trial was initiated by the monkey pressing down on a joystick, which illuminated a fixation square. After fixation acquisition, a white cue-ring was flashed (133ms) in the periphery, followed 500ms later by stimulus onset. The stimuli consisted of two circularly windowed colored "checkerboards". One stimulus appeared in the location previously occupied by the cue-ring (the "cued stimulus"), and the other was presented at an equally eccentric opposing location (the "foil stimulus"). A change in mean stimulus saturation was possible 1–4s after stimulus onset; if the cued stimulus changed, the monkey was required to release the joystick within 150–750 ms, and if the foil stimulus changed, the monkey was required to keep the joystick depressed. Only one stimulus change was possible in each trial. Each block had a ratio of 3 cued : 1 foil change trials, presented in pseudorandom order in each block. Color changes were isoluminant changes in mean saturation and were masked by luminance noise.

Single unit recordings

We recorded from 60 SC neurons in monkey 1, and 79 in monkey 2. The details of procedures used to record extracellular activity of SC neurons using single electrodes may be found in our previous paper[7]. Briefly, once a single unit's waveform was well isolated, visual and memory guided saccade tasks were used to map the receptive field (RF) online, and the covert color-change task was started with either the cued stimulus or the foil stimulus in the unit's RF. Data were collected with a Plexon MAP system (Plexon Inc., Dallas, TX), and putative spike waveforms were sorted offline with Plexon Offline Sorter.

Bilateral SC recordings

In 3 sessions in monkey 1, we recorded the activity of both the right and left SC simultaneously using 24-channel v-probes (50μm spacing between contacts; Plexon Inc., Dallas, TX). In each session, once the probes had been advanced into the intermediate / deep layers of each SC, threshold crossings ($\mu - 3\sigma$ on each channel) were used during a visually guided saccade task to simultaneously map RFs on all 48 contacts online. Stimuli were then placed by hand to maximize coverage of RFs. During offline analysis, we identified and kept only those units that had at least 2/3 of the stimulus inside their RF envelope (border defined by 50% of the maximum response to saccade target onset). Unlike single unit recording sessions, in which cue and foil stimulus were always 180° of elevation apart, we placed the two stimuli at 0° and 200° of elevation to align stimulus placement with RFs; for the same reason, we placed the stimuli at 13° eccentricity, rather than the 9–11° that had been used previously. At the conclusion of each session, the monkey completed 30–50 visually guided and 30–50 memory guided saccades with a target placed at the center of each of the two covert task stimulus locations – these saccade tasks were used offline to categorize SC neurons.

Continuous spike channel data recorded with a Plexon Omniplex D were analyzed offline with Kilosort[21] including manual verification and adjustment steps. Following sorting, single neurons were classified following criteria established previously[7]. Only well isolated units with activity during the memory guided saccade task (allowing them to be categorized)

were retained for further analysis. Of 131 units, 80 were either visual-movement or visual-movement prelude, and were used to estimate correlations, as well as being added to the dataset of 34 single units recorded previously in monkey 1. Importantly, activity-d' computed from simultaneous bilateral recordings (2.68) was comparable to the value from single-electrode recordings (2.59), and performance during bilateral recording sessions was well accounted for by a model using neuronal data from only those sessions (figure S2).

Inactivation

We inactivated by injecting muscimol (0.5μl, 5 μg/μl), a $GABA_A$ channel agonist, in the intermediate/deep layers of SC[10], placing the tip of our injection canula 2.5mm below SC surface. In each inactivation session (n = 8 and n = 10 in monkey 1 and 2, respectively), 300–550 trials were collected "before" injection, and a similar number were collected "during" inactivation (within 30 to 240 minutes after injection). During inactivation, visually guided saccades were used to map the extent of the inactivation area as described previously[11], and the cued stimulus was alternately presented inside and outside this area in successive blocks of trials.

Microstimulation

We followed previously established methods for microstimulation of SC without evoking saccades[8]. In each session (n = 7 and n = 6 in monkey 1 and 2, respectively), the microelectrode tip was placed at 2.5mm below SC surface. Before task data were collected, we first evoked 5–8 saccades by stimulating at 350Hz with biphasic current pulses of 25–30μA, and used their average endpoint to subsequently place task stimuli so that either the cued or the foil stimulus was always overlapping this location. We then lowered both stimulation frequency and current pulse amplitude until saccades were evoked < 50% of the time with fixation point off and < 10% of the time with the fixation point on. Typically, these were achieved with a stimulation frequency of 70–85Hz and a current amplitude of 10–15uA, consistent with previous reports[22]. During the attention task, microstimulation onset was 300ms prior to stimulus change and had a duration of 600ms. There were an equal number of all trial types "with" microstimulation and "without" microstimulation, and these were pseudorandomly intermixed.

SC output simulations

Spike data from visual-movement (n = 75) and visual movement prelude (n = 76) neurons was used to simulate SC output activity (for neuron classification details, see our previous paper[7]). As mentioned in the results, these classes were selected to maximize the model's ability to account for behavior in our task based on previous findings[13], but we cannot exclude the possibility that other classes of SC neurons contribute as well. Spike counts were binned in a sliding 100ms window (1ms increments). Each trial's data comprised a portion of spike rate activity aligned on stimulus onset [−250ms, + 2000ms] concatenated with a portion aligned on stimulus change [−500ms, +1000ms]. In trials where the change occurred in the [+1000ms, +2000ms] interval, activity was truncated at change onset. After binning and concatenation, each neuron's spike count data was normalized by dividing each binned count by the standard deviation of that neuron's counts across trials and conditions. By normalizing activity, each neuron made a roughly equal contribution to performance in

our model; more complex weighting schemes might produce better model performance but would require additional assumptions.

We generated 10000 simulated “cue change on right” and 10000 “foil change on left” trials to estimate the statistics of SC output. Each trial included averaged activity from simulated left and right SC pools of 1000 neurons. For “cue change on right” trials, left SC activity incorporated normalized spike data from trials where the cued change was contralateral to the neuron, and right SC activity was from trials where the cued change was ipsilateral. Similarly, for “foil change on left” trials, right SC activity was from trials where the foil change was contralateral and left SC activity was from trials where the foil change was ipsilateral. A single simulated trial incorporated one trial’s worth of normalized spike count data from each neuron in the pool for left SC activity, and one trial’s worth for right SC activity (in each monkey). For example, a simulated “cue change on right” trial in monkey 1 was built by randomly drawing and averaging 1000 “cue change contralateral” trials and 1000 “cue change ipsilateral” trials (with replacement) from data across 114 neurons. Importantly, for cross-validation, we reserved some trials for training and others for testing (see below).

We estimated the statistics of SC output at each millisecond in a trial and incorporated correlations using established methods[13]. Separately for cue-change and foil-change trials, we estimated the mean and variance of the pooled right and left SC activity across trials: $\mu_R[t]$, $\sigma_R[t]$, $\mu_L[t]$, $\sigma_L[t]$. As described in the results, a trajectory in right SC vs. left SC activity space will cross the decision boundary if its “summary point” (activity at time of maximum difference between right and left SC) lies outside the decision boundary (figure 1e). We therefore focused our simulations of the effects of correlations on SC output at the time of maximum difference between left and right SC activity: t_{MD}. Importantly, this approach – determining model responses based on activity at a single time-point – remains valid after incorporating the effects of correlations as long as the correlations don’t vary substantially in time; we verify this below. We incorporated the effects of within-SC correlations (ρ_w) and between-SC correlations (ρ_b) by constructing a 2n X 2n (n = 1000) covariance matrix with a specific correlation structure. The covariance matrix incorporating correlations (Σ_c) was computed from an uncorrelated covariance (Σ_{uc}) and a correlation matrix (P): $\Sigma_c = \Sigma_{uc}^{\frac{1}{2}} P \Sigma_{uc}^{\frac{1}{2}}$. Where:

$$P = \begin{bmatrix} 1 & \rho_w & \rho_w & & \rho_b & \rho_b & \rho_b \\ \rho_w & 1 & \rho_w & \cdots & \rho_b & \rho_b & \rho_b \\ \rho_w & \rho_w & 1 & & \rho_b & \rho_b & \rho_b \\ \vdots & & & \ddots & & \vdots & \\ \rho_b & \rho_b & \rho_b & & 1 & \rho_w & \rho_w \\ \rho_b & \rho_b & \rho_b & \cdots & \rho_w & 1 & \rho_w \\ \rho_b & \rho_b & \rho_b & & \rho_w & \rho_w & 1 \end{bmatrix}$$

$$\Sigma_c = \begin{bmatrix} n\sigma_R^2 & 0 & 0 & & & 0\,0\,0 \\ 0 & n\sigma_R^2 & 0 & \cdots & & 0\,0\,0 \\ & & & & & 0\,0\,0 \\ 0 & 0 & n\sigma_R^2 & & & \\ \vdots & & \ddots & & \vdots & \\ & & & & n\sigma_L^2 & 0 \quad 0 \\ 0\,0\,0 & & & & & \\ 0\,0\,0 & & \cdots & 0 & n\sigma_L^2 & 0 \\ 0\,0\,0 & & & & & \\ & & & 0 & 0 & n\sigma_L^2 \end{bmatrix}$$

Using identical values of ρ_b and ρ_w throughout P and identical values of σ_R and σ_L throughout Σ_{uc} is appropriate because of our pooling approach: variability in these parameters will be eliminated after averaging. Further, because our model relies only on the pooled (averaged) activity, and not the activity of individual units in the pool, instead of the 2n X 2n covariance, we can directly compose an expression for the 2 X 2 covariance matrix of the "summary points" incorporating correlations:

$$\overline{\Sigma}_c = \begin{bmatrix} \sigma_R^2 + (n-1) * \rho_w * \sigma_R^2 & \sigma_R * \sigma_L * \rho_b \\ \sigma_R * \sigma_L * \rho_b & \sigma_L^2 + (n-1) * \rho_w * \sigma_L^2 \end{bmatrix}$$

To generate "summary points" with correlations incorporated, we drew from a distribution with mean ($\mu_R[t_{MD}]$, $\mu_L[t_{MD}]$) and covariance $\overline{\Sigma}_c$, setting $\rho_w = 0.09$ and $\rho_b = 0.047$, the mean correlation values measured in monkey 2. As mentioned above, parameters were estimated separately for cue change and foil change trials, meaning that generating new "summary points" relied on one mean + covariance for cue change and one mean + covariance for foil change trials (for each monkey).

In simulating the effects of varying n (pool size), ρ_w, and ρ_b on d' (figure 2a-c), the primary constraint is that any valid covariance matrix (such as Σ_c) must be positive-semidefinite (its eigenvalues must be non-negative). From this, it follows that the average correlation among a set of m random variables has lower bound: $\rho > -1/(m-1)$[23]; we considered $\rho_w \in [0, 0.2]$. In all simulations, we computed Σ_c and or $\overline{\Sigma}_c$ by first fixing ρ_w and then finding the range of ρ_b values that keep Σ_c and or $\overline{\Sigma}_c$ positive-semidefinite: $(\rho_w*(n-1) + 1)/n \geq \rho_b \geq -(\rho_w*(n-1) + 1)/n$. Intuitively, ρ_b can be more negative than $-1/(n-1)$ because it is the correlation of a subset (n^2) of all possible $\binom{2n}{2}$ pairs (2n = n right + n left SC neurons).

The local minimum "trough" in figure 2c can be characterized by closed-form expressions for the slope "m" and intercept "b" of the isoclines of d' that can be derived from $\overline{\Sigma}_c$:

$$m = \frac{2 \cdot \left(\sigma^2_{cue_x} + \sigma^2_{cue_y} + \sigma^2_{foil_x} + \sigma^2_{foil_y}\right)}{\sigma_{cue_x} \cdot \sigma_{cue_y} + \sigma_{foil_x} \cdot \sigma_{foil_x}}$$

$$b = \frac{2 \cdot \Delta\mu^2}{d'} \cdot \frac{1}{\sigma_{cue_x} \cdot \sigma_{cue_y} + \sigma_{foil_x} \cdot \sigma_{foil_x}}$$

Using these expressions, the slope and intercept of the trough are (m = 0.8, b = 0.06) in monkey 1, and (0.84, 0.06) in monkey 2. Since the slopes are non-unity the relative difference in correlations ($\rho_w - \rho_b$) is not constant along the trough; this is noteworthy because previous pooling models have been shown to be insensitive to absolute changes in ρ_w and ρ_b, caring only about the relative difference $\rho_w - \rho_b$[24], we are unaware of any examples that have shown the type of sensitivity we here demonstrate.

Correlation measurements

We measured within-SC (ρ_w) and between-SC (ρ_b) correlations at the time during the trial when they would affect activity-d' - when the difference between right and left SC activity was maximal (t_{MD}). In each of the 3 bilateral SC recording sessions, we examined 4 trial conditions: (1) cue-change on left, (2) cue-change on right, (3) foil-change on left, (4) foil-change on right. Stimulus saturation and luminance were i.i.d. across trials and conditions at the time correlations were measured, but the precise values varied from trial to trial (see our previous work for statistical details[13]). Spike counts were binned in a 100ms sliding window (1ms increments), and each neuron's data was normalized by dividing its binned counts by the standard deviation of that neuron's counts across trials and conditions. We restricted our analysis to a portion of activity aligned on stimulus onset [−250ms, + 2000ms] concatenated with a portion aligned on stimulus change [−500ms, +1000ms]. Over the 100ms sliding window, we computed the pairwise correlation between spike counts of each pair of neurons within right or left SC (n = 1084), or between right and left SC (n = 1022) at each millisecond in the trial. However, we also computed correlations at the time of maximum difference between right and left SC in each trial (t^i_{MD}). To find t^i_{MD}, we computed the difference of the average normalized activity of all right SC neurons and all left SC neurons in each trial and identified when that difference was maximal. For each neuron, we extracted the normalized count at t^i_{MD} in each trial and then computed the trial-by-trial Pearson correlation between pairs of neurons.

Next we sought to determine whether there was significant variation in ρ_w (1) over time, (2) depending on trial condition, or (3) depending on side (i.e. contralateral or ipsilateral to the stimulus change). We performed a repeated measures ANOVA (rmANOVA) on ρ_w values in a window surrounding the average time of maximal difference between left and right SC activity ($\bar{t}_{MD}$). We included correlation values for each time bin from $\bar{t}_{MD}$ - 150ms to $\bar{t}_{MD}$ + 150ms, and also included the ρ_w values t^i_{MD} as an additional time-point. The factors

included were trial condition (cue change or foil change), and side (whether the ρ_w value was from a pair within the right SC or the left SC). We found no significant variation over time ($p = 0.42$, $df = 300$, $F = 1.01$), depending on trial condition ($p = 0.9$, $df = 900$, $F = 0.89$), or depending on side ($p = 0.62$, $df = 300$, $F = 0.97$). A similar rmANOVA on ρ_b values (with only the trial condition factor since between-SC pairs have no "side" per se) also revealed no significant variation depending on time ($p = 0.42$, $df = 300$, $F = 1.01$) or trial condition ($p = 0.61$, $df = 300$, $F = 0.98$). The lack of variation in ρ_w and ρ_b over time validates our approach to incorporating the effects of measured correlations into our model (i.e. applying measured correlations to "summary points"). The lack of variation in correlation values over trial condition, side, or time, suggests that it is appropriate to incorporate the same ρ_w and ρ_b values in all conditions.

Cross Validation

We cross validated our model's predictions by using a training set of spike count data to fit the model's decision boundary and a test set of data to calculate the model's hit and false alarm rate predictions. More specifically, we used the training set to generate "summary points" (n = 10000) incorporating correlations as described above and fit the decision boundary using these (see below for details of boundary fitting). We then generated a new set of "testing summary points" (n = 10000) using the reserved test set of trials and used the fitted boundary to compute a hit and false alarm rate.

To parse the data into training and test sets of simulated trials, we divided each neuron's trials equally within each trial type (cue change on left, cue change on right, foil change on left, foil change on right). After dividing, in each monkey, there were > 1000 trials in each trial-type-grouping, across neurons.

Model fitting

We fit the model to hit and false alarm rates from each recording session (n = 59 in monkey 1 and n = 70 in monkey 2), each inactivation session (n = 8 in monkey 1, n = 10 in monkey 2), and each microstimulation session (n = 7 in monkey 1, n = 6 in monkey 2). The model had 1 parameter (the intercept) that defined a pair of boundaries that were mirror-symmetric across the identity line (figure 1e): one boundary was defined by the line "y = x + a" and the other by "y = x – a", where "a" is the fitted parameter. The parameter was estimated by minimizing a cost function in MATLAB (The Mathworks, Inc.): the sum of the negative log likelihoods of the model-predicted hit and false alarm rates (h_a and f_a), given the monkey's hit and false alarm rates (h_m and f_m):

$$-\left[\log P\left(h_a \middle| h_m\right) + \log P\left(f_a \middle| f_m\right)\right] \text{ where } P(x|y) = \binom{n}{k} * x * (1-x)^{n-k}, y = \frac{k}{n}.$$

Fitting inactivation and microstimulation session performance additionally incorporated an additional parameter that scaled SC activity, and reflected the structure of those experimental sessions. For inactivation sessions the boundary was fit to "before" hit and false alarm rates, and was then fixed while the scaling parameter was fit to "during" rates. For microstimulation sessions (in which trials "with" and "without" stimulation were

interleaved), the boundary was fit simultaneously using hit and false alarm rates from "with" and "without' stimulation trials. All fitting again used the "training" set of reconstructed trials and relied on minimizing negative log likelihood, with the multiplicative scaling parameter initialized at 1 and the additive scaling parameter (when used) initialized at 0.

To determine whether activity scaling alone suffices to account for performance during inactivation / with microstimulation, we fit a model that featured both scaling and 2 separate decision boundaries (1 before inactivation / without microstimulation, and 1 during inactivation / with microstimulation). We fit this 3 parameter model (1 scaling value, 2 decision boundary intercepts) by first fitting 1 decision boundary to a single session's before inactivation / without microstimulation hit and false alarm rate, then simultaneously fitting 1 scaling value and 1 decision boundary to during the inactivation / with microstimulation hit and false alarm rate. We quantified the performance of each model by computing absolute prediction errors for hit and false alarm rates in the perturbation conditions (during inactivation or with microstimulation) because in the no-perturbation conditions there is no scaling.

Statistics

To compare model performance to monkey performance during recording sessions (figure 1f), we used a generalized linear model (GLM) using MATLAB. All monkey and model hit and false alarm rates (response counts and trial counts) were combined into a single response variable (*rates*; n = 504 observations). There were 3 categorical predictors: (1) monkey or model (*MM*), (2) hit or false alarm (*HF*), (3) monkey 1 or monkey 2 (*M12*). Responses were assumed to follow a binomial distribution (though this was not formally tested), and a logistic link function was used. In Wilkinson notation, the model formula was: logit(*rates*) ~ 1 + *MM*HF*M12* meaning that the model included a constant term to capture bias, individual terms for each of the predictors, as well as terms for interactions amongst the predictors (m = 497 error degrees of freedom).

To compare models accounting for performance changes during inactivation or microstimulation, we used a 3 factor ANOVA: (1) monkey (1 or 2), (2) model (multiplicative scaling, additive scaling, or multiplicative scaling with variable boundary), and (3) perturbation condition (inactivation or microstimulation). Data were hit and false alarm rate errors (monkey – model) during inactivation or with microstimulation only. There were 185 degrees of freedom, and 172 error degrees of freedom. Only the "model" factor and "model:condition" interaction terms were significant ($p << 0.01$; model df = 2, F = 11.76, model:condition df = 2, F = 14.76). Post-hoc Tukey-Kramer testing with $\alpha = 0.05$ showed that the additive model had significantly larger error than the multiplicative models for inactivation data; for microstimulation data, the errors were equivalent. The same post-hoc testing showed that the multiplicative model with variable boundary had statistically indistinguishable errors compared to the multiplicative model with constant boundary.

To test whether model success depended on a unique relationship between monkey performance and its own SC activity we compared the means and variances of model errors in the default "unswapped" case (using monkey 1's SC activity to predict monkey 1's inactivation / microstimulation performance, and using monkey 2's activity to predict

monkey 2's performance), to the "swapped" case (using monkey 1's SC activity to predict monkey 2's performance and vice-versa) by bootstrapping. Model errors were the absolute value of the difference between the model predicted rate and the monkey's actual rate (e.g. | monkey 1 hit rate - model predicted hit rate|). To test whether the "unswapped" and "swapped" model errors had the same or different means / variances, we pooled "unswapped" and "swapped" errors, resampled 10000 times with replacement to generate a test statistic distribution, and compared the actual value (difference between "unswapped" and "swapped") to the bootstrapped distribution to obtain a p-value.

One-sample t-tests on measured between-SC and within-SC correlation values tested the null hypothesis that those values were normally distributed with zero mean and unknown variance. Between-SC values t-statistic = 8.8621, df = 1083; within-SC values t-statistic = 14.1342, df = 1021.

The details of a repeated measures ANOVA (rmANOVA) on measured correlations can be found in the Correlations subsection of the Methods section.

No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications[25,26]. The use of t-tests and ANOVAs assumes data distributions were normal but this was not formally tested.

Life Sciences Reporting Summary.

Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability

The code used in the current study is available from the corresponding author(s) on reasonable request.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author(s) on reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank B. Cumming, M.Shadlen, and M.Leathers for helpful comments on a previous draft of this manuscript. We thank F.Arcizet, A.Bollimunta, A.Bogadhi, L.Wang, and C.Quaia for helpful discussions. This work was supported by the National Eye Institute Intramural Research Program at the National Institutes of Health.

Main Text References

1. Ratcliff R & McKoon G Neural Comput 20, 873–922 (2008). [PubMed: 18085991]
2. Gold JI & Shadlen MN Annu. Rev. Neurosci. 30, 535–574 (2007). [PubMed: 17600525]
3. Hanks TD, Ditterich J & Shadlen MN Nat Neurosci 9, 682–689 (2006). [PubMed: 16604069]

4. Katz LN, Yates JL, Pillow JW & Huk AC Nature 535, 285–288 (2016). [PubMed: 27376476]
5. Basso MA & Wurtz RH Journal of Neuroscience 18, 7519–7534 (1998). [PubMed: 9736670]
6. White BJ et al. Nature Communications 8, 14263 (2017).
7. Herman JP & Krauzlis RJ eNeuro 4, ENEURO.0046–17.2017 (2017).
8. Cavanaugh J, Alvarez BD & Wurtz RH J. Neurosci. 26, 11347–11358 (2006). [PubMed: 17079663]
9. Müller JR, Philiastides MG & Newsome WT Proc Natl Acad Sci USA 102, 524–529 (2005). [PubMed: 15601760]
10. Lovejoy LP & Krauzlis RJ Nat Neurosci 13, 261–266 (2010). [PubMed: 20023651]
11. Zénon A & Krauzlis RJ Nature 489, 434–437 (2012). [PubMed: 22972195]
12. Green DM & Swets JA (Wiley: 1966).
13. Shadlen MN, Britten KH, Newsome WT & Movshon JA Journal of Neuroscience 16, 1486–1510 (1996). [PubMed: 8778300]
14. Averbeck BB & Lee DJ Neurophysiol. 95, 3633–3644 (2006).
15. Moreno-Bote R et al. Nat Neurosci 17, 1410–1417 (2014). [PubMed: 25195105]
16. Abbott LF & Dayan P Neural Comput 11, 91–101 (1999). [PubMed: 9950724]
17. Cohen MR & Kohn A Nat Neurosci 14, 811–819 (2011). [PubMed: 21709677]
18. Reynolds JH & Heeger DJ Neuron 61, 168–185 (2009). [PubMed: 19186161]
19. Lovejoy LP & Krauzlis RJ Proc. Natl. Acad. Sci. U.S.A. 114, 6122–6126 (2017). [PubMed: 28533384]
20. Sridharan D, Steinmetz NA, Moore T & Knudsen EI Journal of Neuroscience 37, 480–511 (2017). [PubMed: 28100734]
21. Pachitariu M, Steinmetz N, Kadir S, Carandini M & Harris KD bioRxiv (2016).doi: 10.1101/061481
22. Glimcher PW & Sparks DL J. Neurophysiol. 69, 953–964 (1993). [PubMed: 8463820]
23. Helstrom CW (Pergamon: 1975).
24. Nienborg H, Cohen MR & Cumming BG Annu. Rev. Neurosci. 35, 463–483 (2012). [PubMed: 22483043]
25. Bondy AG, Haefner RM & Cumming BG Nat Neurosci 79, 1–15 (2018).
26. Nevet A, Morris G, Saban G, Arkadir D & Bergman HJ Neurophysiol. 98, 2232–2243 (2007).

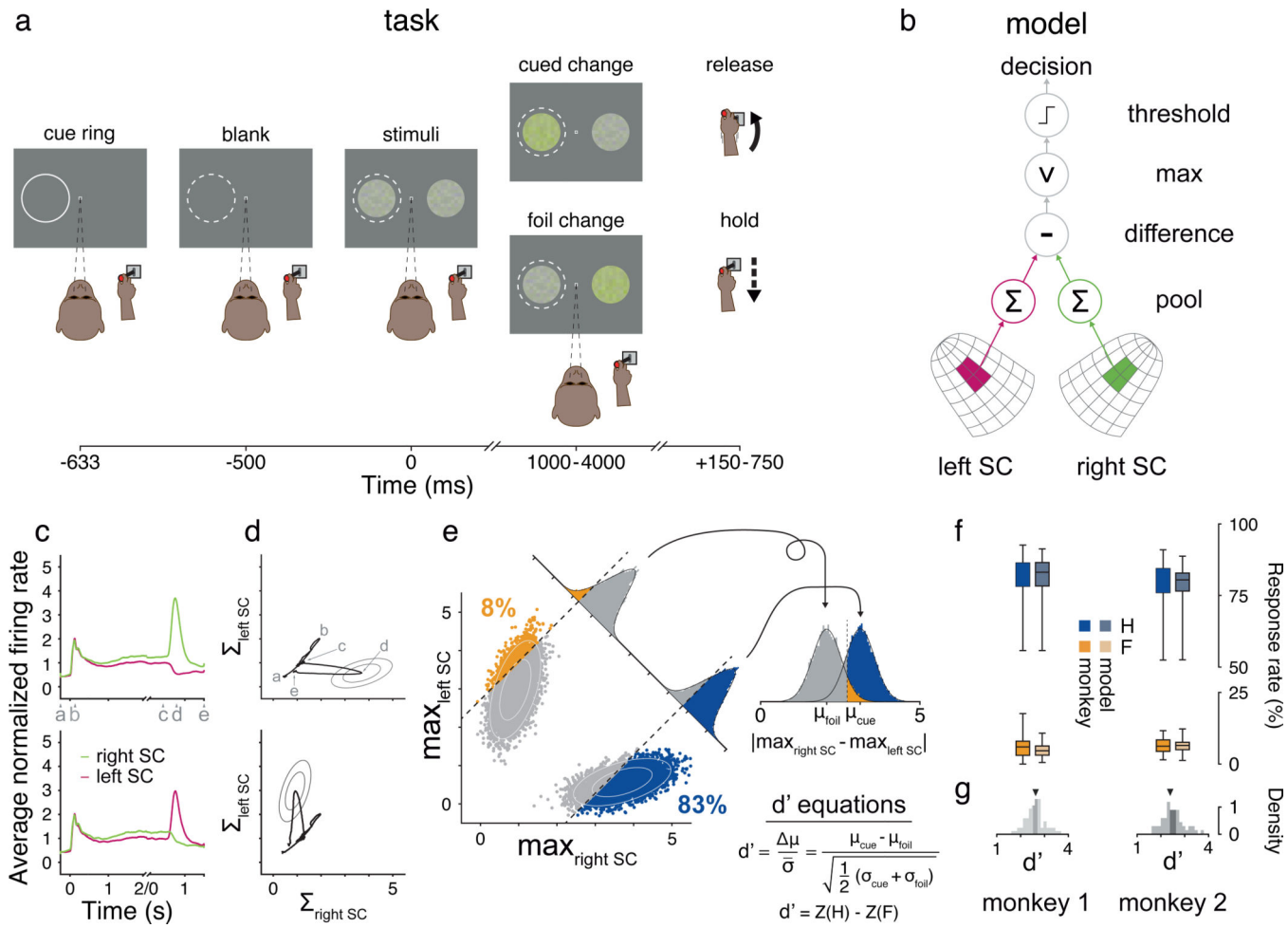

Figure 1.

Task, model, and performance predictions. (a) The monkey was required to maintain central fixation and hold down a joystick throughout each trial, releasing the joystick to indicate that the cued stimulus had changed. Only one stimulus changed per trial. Dashed white cue-ring was not visible to the monkey. (b) Single-trial activity from simulated pools of right and left SC neurons (n = 1000 each) with RFs overlapping one stimulus location were averaged. If the maximum difference between pools at any time in the trial was above a threshold, a decision to respond was triggered. (c) Pooled right (green) and left (red) SC activity averaged across 10000 simulated cue change (top) and foil change trials (bottom); each trial included a portion of activity aligned with respect to stimulus onset (−0.25s to 2s) concatenated with a portion aligned to stimulus change (−0.5s to 1.5s); the tick labeled 2/0 marks where the portions were concatenated. (d) A trajectory in "SC activity space" is constructed by plotting right SC activity-versus-time against left SC. Characters (a,b,c,d,e) with pointers indicate time points corresponding to matching characters along time axis in panel c. Dark gray ellipses surround the point on the average trajectory when the difference between right and left SC activity is maximal (point d), and indicate the 68% and 95% confidence intervals on that maximum difference. (e) Relationship between 2D activity space and 1D activity difference representations. In 2D space (left), each "summary point" is

the point on the trajectory at the time of maximum difference between right and left SC activity in one simulated trial. Ellipses are identical to those plotted in panel d. Dashed black lines are a pair of decision of boundaries (1 parameter) fit to monkey 1's hit (83%) and false alarm rate (7%) from an example session. Blue summary points are from simulated cue change trials that crossed the decision boundary (n = 8283) and are counted as "hits" while orange summary points (n = 757) are "false alarms". In 1D (inset), model hit and false alarm rates can equivalently be calculated by projecting each summary point onto the line y = -x (a simple difference of the x- and y-value), and counting the number that exceed the criterion corresponding to the 2D decision boundary. Taking the absolute value of the 1D projection "flips" the 1D foil distribution such that criterion values align (right), resembling a familiar signal detection theory presentation. In this format, an *activity-d'* (computed from means and standard deviations of the cue and foil distributions: μ_{cue}, σ_{cue} and μ_{foil}, σ_{foil}) can be compared to a *behavioral-d'* (computed from monkey hit (H) and false alarm (F) rates). As a complement to these 1D and 2D representations, figure S1 illustrates the temporal distribution of summary points. (f) Box plot summarizing hit (blue) and false alarm rates (orange) for monkey (high saturation fills) and model (low saturation fills; n = 59 in monkey 1, n = 70 in monkey 2). Upper and lower box edges mark 25th and 75th percentile of distribution, whiskers mark 2.5th and 97.5th percentiles, and central line marks median (50th percentile). (g) Distributions of behavioral-d' values computed from each session's hit and false alarm rates (n = 59 in monkey 1, n = 70 in monkey 2). Dark gray shading indicates bootstrapped 95% confidence interval on the mean of each distribution (monkey 1: [2.37, 2.63]; monkey 2: [2.29, 2.48]). Black triangle indicates each monkey's SC activity-d' (2.61 and 2.39).

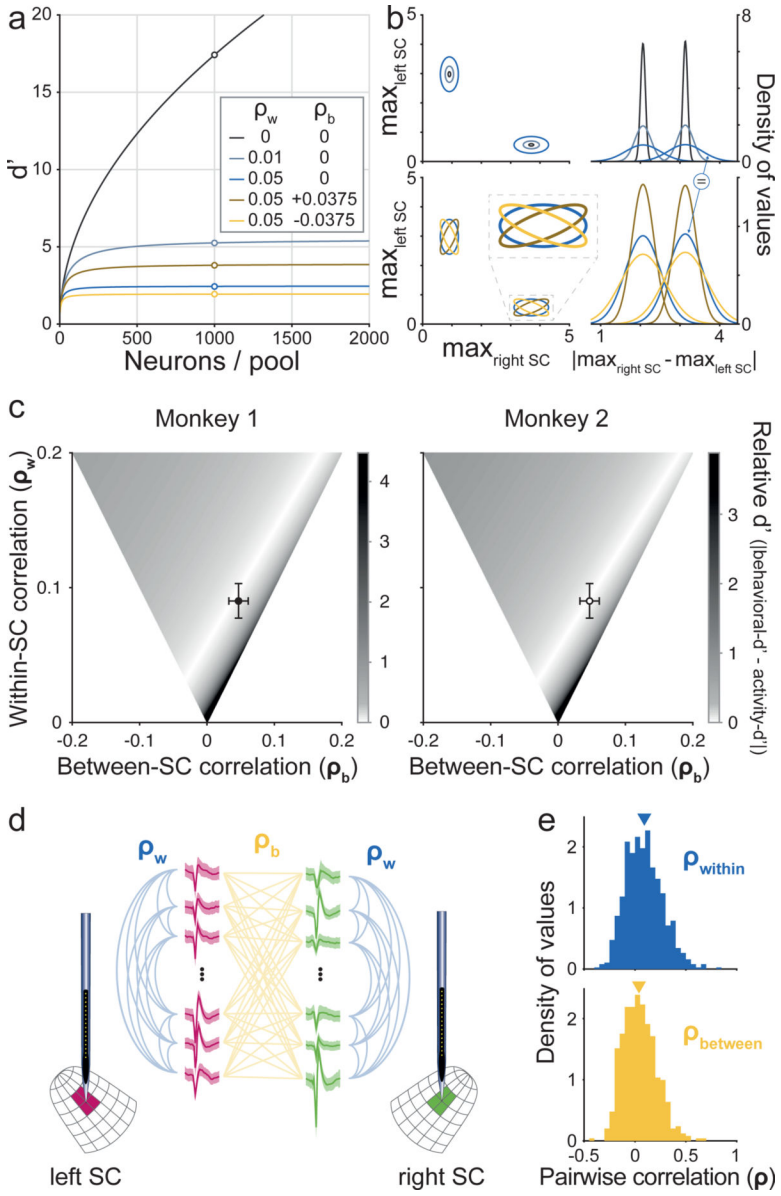

Figure 2.

Correlation effects and measurements. (a) SC activity-d' as a function of pool size (neurons per pool) for combinations of average correlation within (ρ_w) and between (ρ_b) pools. White-filled circles correspond to example cases illustrated in panel b. (b) In fixed pools of 1000 neurons, increasing ρ_w increases x- and y-components of 2D covariance (left panel) and 1D variance (right panel). Right, bottom: $\rho_b > 0$ (brown) "tilts" 2D distributions towards identity line (left), reducing 1D variance (right); $\rho_b < 0$ (yellow) tilts 2D distributions away from identity line (left) increasing 1D variance (right). (c) Relative d' (magnitude of difference: activity-d' – behavioral-d') with fixed pool size (n = 1000 each), for combinations of ρ_w and ρ_b. ρ_w was systematically varied over 0.0–0.2; for each ρ_w value, ρ_b was varied over the range of possible theoretical values (see methods). Overlaid symbols (solid black and black-outlined) are from measurements in monkey 1, and are identical in

left and right panels, indicating the mean and 95% confidence interval on measured ρ_w and ρ_b values. (d) ρ_w (blue) and ρ_b (yellow) were measured in monkey 1 by simultaneous bilateral multi-contact probe recordings. Colored connecting lines represent hypothetical neuronal pairs used to measure within-SC and between-SCs correlations. (e) Distributions of ρ_w (blue) and ρ_b (yellow) values measured at "summary point" time (maximum difference between x- and y-component of 2D trajectory); colored triangles indicate means ($\rho_w = 0.09$ and $\rho_b = 0.047$).

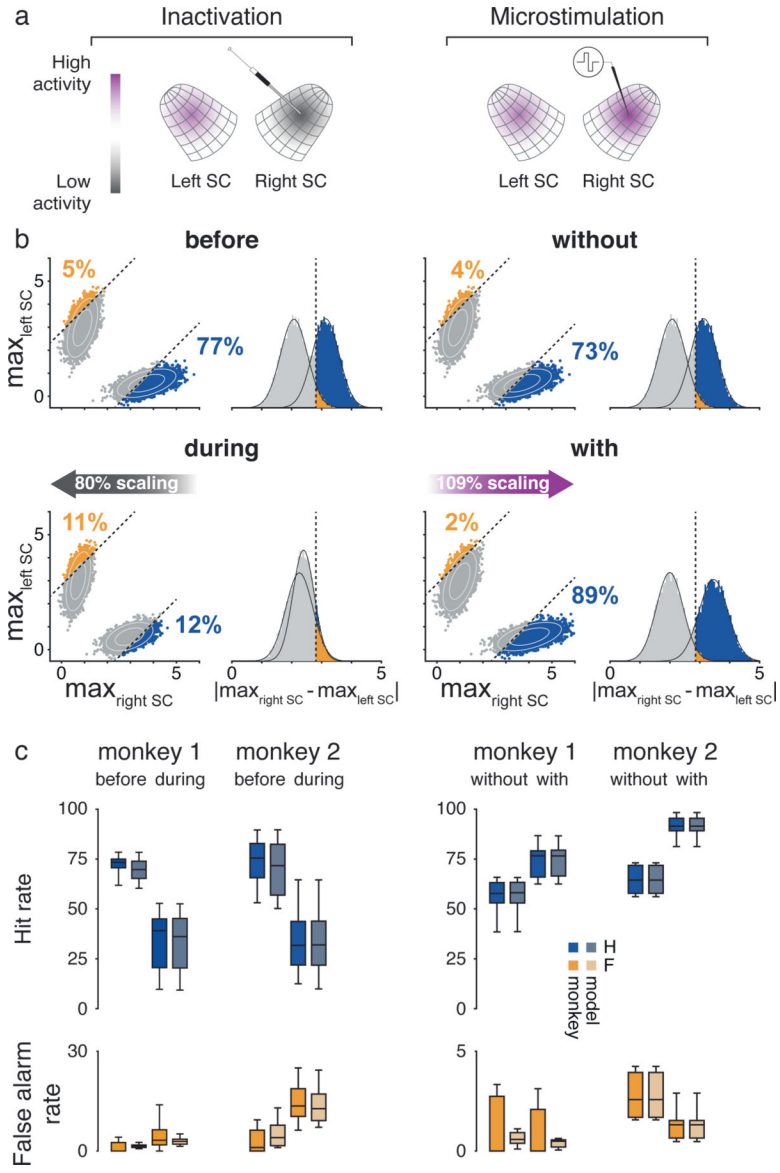

Figure 3.

Modeling the effects of unilateral causal manipulations of SC activity. (a) Following protocols known to affect behavior, in separate sessions we either damped activity by injecting the $GABA_A$ agonist muscimol (left) or facilitated activity by microstimulation (right). (b) We simulated the effects of these causal manipulations by scaling SC activity. Top: Example decision boundary fits to hit and false alarm rates for the control conditions: before muscimol (left), or without microstimulation (right), without scaling of SC activity; format as in figure 1e. Bottom: with the decision boundary held fixed, the scaling parameter was fit to hit and false alarm rates during muscimol (left) or microstimulation (right). (c) Monkey and model performance before and during inactivation (left; n = 8 in monkey 1, n = 10 in monkey 2) or without and with microstimulation (right; n = 7 in monkey 1, n = 6 in monkey 2); box plots and other conventions are as described in figure 1f.