

UCSF

UC San Francisco Previously Published Works

Title

Neural Representation and Causal Models in Motor Cortex

Permalink

<https://escholarship.org/uc/item/20k2j560>

Journal

Journal of Neuroscience, 37(12)

ISSN

0270-6474

Authors

Chaisanguanthum, Kris S
Shen, Helen H
Sabes, Philip N

Publication Date

2017-03-22

DOI

10.1523/jneurosci.1000-16.2017

Peer reviewed

Neural Representation and Causal Models in Motor Cortex

Kris S. Chaisanguanthum,^{1,2,3} Helen H. Shen,^{1,2} and Philip N. Sabes^{1,2,3}

¹Department of Physiology, ²Center for Integrative Neuroscience, and ³Sloan-Swartz Center for Theoretical Neurobiology, University of California, San Francisco, San Francisco, California 94143-0444

Dorsal premotor (PMd) and primary motor (M1) cortices play a central role in mapping sensation to movement. Many studies of these areas have focused on correlation-based tuning curves relating neural activity to task or movement parameters, but the link between tuning and movement generation is unclear. We recorded motor preparatory activity from populations of neurons in PMd/M1 as macaque monkeys performed a visually guided reaching task and show that tuning curves for sensory inputs (reach target direction) and motor outputs (initial movement direction) are not typically aligned. We then used a simple, causal model to determine the expected relationship between sensory and motor tuning. The model shows that movement variability is minimized when output neurons (those that directly drive movement) have target and movement tuning that are linearly related across targets and cells. In contrast, for neurons that only affect movement via projections to output neurons, the relationship between target and movement tuning is determined by the pattern of projections to output neurons and may even be uncorrelated, as was observed for the PMd/M1 population as a whole. We therefore determined the relationship between target and movement tuning for subpopulations of cells defined by the temporal duration of their spike waveforms, which may distinguish cell types. We found a strong correlation between target and movement tuning for only a subpopulation of neurons with intermediate spike durations (trough-to-peak $\sim 350 \mu\text{s}$ after high-pass filtering), suggesting that these cells have the most direct role in driving motor output.

Key words: computational model; electrophysiology; macaque; motor cortex; reaching; variability

Significance Statement

This study focuses on how macaque premotor and primary motor cortices transform sensory inputs into motor outputs. We develop empirical and theoretical links between causal models of this transformation and more traditional, correlation-based “tuning curve” analyses. Contrary to common assumptions, we show that sensory and motor tuning curves for pre-movement preparatory activity do not generally align. Using a simple causal model, we show that tuning-curve alignment is only expected for output neurons that drive movement. Finally, we identify a physiologically defined subpopulation of neurons with strong tuning-curve alignment, suggesting that it contains a high concentration of output cells. This study demonstrates how analysis of movement variability, combined with simple causal models, can uncover the circuit structure of sensorimotor transformations.

Introduction

A central question in the study of voluntary movement is how the activity of cortical neurons effects the transition from motor intention to motor output. Premotor and primary motor (M1) cortices play a central role in this process, and historically most

studies of these areas have taken a representational approach, asking whether and how the activity of individual neurons are “tuned” to various task or movement parameters (Georgopoulos et al., 1982; Wise et al., 1983; Scott and Kalaska, 1995; Crammond and Kalaska, 1996; Cisek et al., 2003; Paninski et al., 2004). This work has demonstrated that the activity of motor/premotor neurons contains information about a range of parameters such as target location, movement velocity, and endpoint force. This representational approach is limited, however, in that these empirical descriptions of neural activity do not identify the mechanisms by which activity tuning arises nor how the underlying neural circuit drives movement. Other, more recent studies have focused on the causal role of premotor and motor cortices, arguing that motor cortical activity is better understood as driving movement or muscle activation, either directly (Todorov, 2000) or

Received March 17, 2016; revised Jan. 31, 2017; accepted Feb. 5, 2017.

Author contributions: K.S.C., H.H.S., and P.N.S. designed research; K.S.C. and H.H.S. performed research; K.S.C., H.H.S., and P.N.S. analyzed data; K.S.C. and P.N.S. wrote the paper.

This work was supported by the Swartz Foundation, the National Eye Institute, and the National Institute of Mental Health—National Institutes of Health (Grants R01EY01569 and P50MH077970).

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Philip N. Sabes, Department of Physiology, University of California, San Francisco, 675 Nelson Rising Lane, San Francisco, CA 94143-0444. E-mail: sabes@phy.ucsf.edu.

DOI:10.1523/JNEUROSCI.1000-16.2017

Copyright © 2017 the authors 0270-6474/17/373413-12\$15.00/0

through induced neural dynamics (Churchland et al., 2010; Lillicrap and Scott, 2013; Shenoy et al., 2013), rather than in terms of movement parameter tuning.

Here, we try to link these two approaches, analyzing tuning of a population of neurons in the context of a causal model. By “causality,” we refer to the flow of information through the neural circuit: we explicitly model the neural population in dorsal premotor (PMd) cortex/M1 as part of the causal chain between stimulus and behavior. Using a center-out reaching task, we measure two experimentally accessible “tuning” functions: how the preparatory activity of a neuron covaries with the task-relevant input to PMd/M1 (in this case the location of the reach target) and how that activity covaries with the output of the circuit (here the initial movement velocity). Whereas these two tuning curves are often implicitly assumed to be the same, e.g., when using target tuning to predict movement velocity (Georgopoulos et al., 1982, 1988; Schwartz and Moran, 2000), we find that this is generally not the case.

We propose that the observed relationship between target and movement tuning curves reflects how the underlying neural circuit implements the sensorimotor transformation. We therefore build a simple linear model of that transformation, using a normative assumption that the model parameters are chosen to minimize movement variability. The linear model as constructed is directional, and thus causal, in the sense that information (both signal and noise) moves only in one direction. We then compare the predictions of the model with the empirical data. This comparison was performed for the whole population and for subpopulations of neurons defined in terms of their spike waveform durations, which have been used previously to identify cell types from extracellular recordings in macaque (Mountcastle et al., 1969; Mitchell et al., 2007; Merchant et al., 2008; Cohen et al., 2009; Kaufman et al., 2010; Song and McPeck, 2010; Yokoi and Komatsu, 2010).

This work provides a theoretical and empirical link between correlation-based tuning-curve analyses of motor cortex and causal models. Furthermore, it shows how simple, model-based analyses of activity in motor cortical neurons can begin to uncover the circuit structure of those neurons and functional roles that groups of neurons play in generating behavior.

Materials and Methods

Two male rhesus macaque monkeys were used in this experiment. All procedures were approved by the Institutional Animal Care and Use Committee of the University of California, San Francisco, and followed NIH guidelines for care and treatment of laboratory animals.

Behavioral task. The monkeys were trained to make planar reaching movement to visual targets with an instructed delay. Data were then collected in 12 sessions for Monkey D and in 14 sessions for Monkey E. The experimental data used in this study were also used in a previous study (Chaisanguanthum et al., 2014), and additional experimental details can be found there. Briefly, reach targets (7 cm from initial hand position) and visual feedback were provided via a virtual reality setup (McGuire and Sabes, 2011). For each trial, animals moved their hand to a fixed “start” position, and a reach target was then presented. After an instructed delay period (drawn uniformly between 800 and 1300 ms), an audible “go” tone was presented, at which point the animal began movement; visual feedback of the arm was extinguished at movement onset. Eye position was monitored, and animals were required to maintain

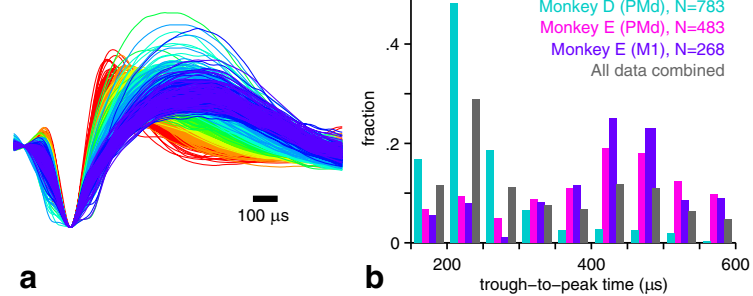


Figure 1. Spike waveform analysis. *a*, Mean spike waveforms, colored according to waveform trough-to-peak time. Colors are for presentational clarity. *b*, Histogram of fraction of cells recorded in each spike waveform width bin. Spike-width distributions vary across microarrays, likely attributable to implantation depth. Note that because the electrophysiological signals were high-pass filtered, the waveform durations reported here are shorter than the true durations.

visual fixation of the target from the initial target presentation until the end of the movement. Fluid reward was delivered in a graded manner based on final endpoint error, and trials with endpoint errors >2.5 cm were not rewarded. The average endpoint error was ~ 5 mm from the center of the target. Our movement metric was initial velocity of the reach \vec{y} , defined as the instantaneous velocity of the hand when its magnitude first exceeds 40% of the peak speed for the trial. Here, we specifically focus on the initial movement angle θ_y , i.e., the angle of the vector \vec{y} . This metric was chosen since it reflects premovement motor planning at a point too early to be influenced by sensorimotor feedback of the movement.

For each experimental session, data were collected in two types of trial blocks, which were interleaved. In “eight-target” blocks, five trials were completed with each of eight reach targets spaced uniformly around the azimuth (0, 45, 90, 135°, etc.). The trials were randomly interleaved, with 40 trials per block and a total of ~ 200 trials per session. Data from the eight-target blocks were used to measure the tuning of neurons to the cued target direction. In “three-target” blocks, 70 trials were completed with each of three targets in a single Cartesian quadrant (e.g., 90, 135, and 180°). The trials were randomly interleaved, with 210 trials per block and a total of ~ 900 reaches per session. The large number of repetitions for each target in three-target blocks, combined with natural movement variability (median SD, across target and sessions, of movement direction for each target was 19 and 10° for Monkeys D and E, respectively), allowed us to study the relationship between that movement variability and variability in the underlying neural response. The selected Cartesian quadrant varied from session to session.

Electrophysiological recordings and analysis. Extracellular signals were recorded from the PMd of both animals and also from M1 of Monkey E using chronically implanted 96-microelectrode arrays (Blackrock Microsystems). Single units were identified by spike sorting with the Plexon Offline Sorter, using a combination of automatic (*t*-Dist E-M algorithm; Shoham et al., 2003) and manual processing. Across recordings from 12 experimental sessions for Monkey D and 15 sessions for Monkey E, we identified 1560 well isolated units, though some of those are likely to correspond to the same neurons.

Five trial epochs are defined for firing rate analysis: (1) the pretrial epoch (500 ms before each trial began), (2) the trial start (variable interval from trial start until the animal moves the arm to the start position), (3) the trial-ready epoch (a fixed 300 ms interval between animal attaining the start position and presentation of the reach target), (4) the delay period (variable interval of 500–1000 ms between the target presentation and the go cue), and (5) the peri-movement period (variable interval from the go cue until the time of peak speed). Most of the study focuses on the delay-period activity, since this activity reflects the feedforward movement plan and this period can be assumed if not otherwise noted.

For later analyses, we will study the activity of groups of cells sorted by the shape of their waveforms. The grouping was motivated by the idea that spike waveform shapes can distinguish some cortical cell types (Mountcastle et al., 1969; Mitchell et al., 2007). Specifically, we sampled filtered waveforms at 30 kHz, computed the mean waveform for each unit, and characterized them by

Table 1. Summary of key notations used, in alphabetical order by symbol

Symbol	Meaning	Remarks
$\square_{in}, \square_{out}$	Subpopulation variables	Analogous to non-subscripted variables, for models with multiple subpopulations (see Fig. 6a)
B	Motor weighting matrix	Describes true mapping from firing rate to movement, e.g., implemented through synaptic weights
d_i	Target tuning slope	$\partial R_i / \partial \theta_x$, as implied by \tilde{D}_i
D , \tilde{D}_i , θ_{Di}	Target tuning matrix, row vector, and preferred direction	Describes firing rates as tuned to target \tilde{x}
E	Neural noise covariance	
f_i	Movement tuning slope	$\partial R_i / \partial \theta_y$, either as implied by \tilde{F}_i or via single-neuron regressions against movement direction
F , \tilde{F}_i , θ_{Fi}	Movement tuning matrix, row vector, and preferred direction	Describes firing rates as though cosine tuned to movement \tilde{y}
$N(\mu, \Sigma)$	Normally distributed random variable	Mean μ , (co)variance Σ
Q	Linear decoder coefficients	Solution to multivariate regression $\tilde{y}_0 - \tilde{y} = \mathbf{Q}^T(\mathbf{R} - \mathbf{R}_0)$. It is not generally true that $\mathbf{Q} = \mathbf{B}$ or $\mathbf{Q} = \mathbf{F}$. (See text for details.)
R , R_i	Firing rate	Bold variables denote full population written as vectors.
R ₀ , \tilde{y}_0	Mean neural and behavioral responses to presentation of target \tilde{x}_0	
W	Subpopulation connectivity matrix	Describes connectivity across subpopulations in multiple subpopulation models
\tilde{x} , θ_x , \tilde{x}_0	Target location, direction (Initial) movement velocity, direction	\tilde{x}_0 denotes a specific target
\tilde{y} , θ_y	(Initial) movement velocity, direction	
$\Sigma_{downstr}$	Downstream motor noise	Contribution to Σ_{tot} independent of measured neural population R_i
Σ_{pop}	Population motor noise	Contribution to Σ_{tot} from neural population R_i
Σ_{tot}	Motor noise covariance	Total covariance in \tilde{y} for a fixed target \tilde{x}_0

the length of time between trough and the peak of the waveform. We use cubic-spline interpolation to calculate this interval. Waveform shapes and their distributions are shown in Figure 1.

Reach and movement tuning and tuning slopes. In Results, we model firing rates as a linear function of a two-dimensional vector, either the target location \tilde{x} or, for a given target \tilde{x}_0 , the initial movement velocity, $\tilde{y} - \tilde{y}_0$, where \tilde{y}_0 is the mean initial movement velocity for reaches to that target. However, for simplicity and to increase statistical power, we experimentally manipulated only the target direction and not the target distance or movement speed. When fitting these tuning models to the data, then, we only want to compute the dependence of firing rate on the directional component of these vectors, which we do as follows.

When describing an individual cell's tuning to target location, the linear tuning model is equivalent to cosine tuning in target angle θ_x :

$$R_i(\tilde{x}) = \tilde{D}_i \cdot \tilde{x} + N(0, \sigma_i^2) = |\tilde{D}_i| |\tilde{x}| \cos(\theta_{Di} - \theta_x) + N(0, \sigma_i^2), \quad (1)$$

where θ_{Di} , the angle of vector \tilde{D}_i , is the “preferred target direction” of cell i . We fit the cosine tuning curves of Equation 1 with linear regression on data from the eight-target blocks. From these fits, we compute the slope d_i of the target tuning curve at target \tilde{x}_0 :

$$d_i = \left. \frac{\partial R_i}{\partial \theta_x} \right|_{\tilde{x}=\tilde{x}_0} = -|\tilde{D}_i| |\tilde{x}_0| \sin(\theta_x - \theta_{Di}). \quad (2)$$

Similarly, for repeated movements to target \tilde{x}_0 , we define the slope of a neuron's tuning to initial movement angle θ_y :

$$f_i = \left. \frac{\partial R_i}{\partial \theta_y} \right|_{\tilde{y}=\tilde{y}_0} = -|\tilde{F}_i| |\tilde{y}_0| \sin(\theta_y - \theta_{Fi}), \quad (3)$$

where θ_{Fi} , the angle of vector \tilde{F}_i , is the “preferred movement direction” of cell i . Because the three-target blocks include many repeated trials to a single target, we can take advantage of the natural variability in initial angle \tilde{y} across trials to fit these movement tuning slopes directly to the data using linear regression.

A summary of key variables used is found in Table 1.

The main focus of analysis in the results is the relationship between the target and movement tuning slopes, d_i and f_i , across cells and across targets in the three-target blocks, recognizing that these represent only first-order approximations to the true response function of a cell. However, a direct comparison of these two tuning curves is potentially complicated by the fact that the initial movement angle is not generally an isotropic function of the target angle; rather, factors such as the inertia of the arm lead to anisotropies in initial angle, which are well modeled by an

affine transformation, $\tilde{y} = \mathbf{A}\tilde{x} + \tilde{\beta}$ (Gordon et al., 1994). Therefore, as a conservative preprocessing step (to maximize the expected correspondence between the two tuning curves), we removed the effect of this anisotropy by applying an affine transformation to the initial angles: for each recording session, we used linear regression to fit \mathbf{A} and $\tilde{\beta}$ to the data from eight-target blocks. For simplicity of presentation, analyses of the movement vector used these transformed values (although none of the conclusions depend on this transformation).

Given that the study focuses on a comparison of linear fits performed on trials from two different trials blocks, we need to verify that the relationship between neural activity and behavior does not change between the three-target blocks and eight-target blocks. These blocks were interleaved (typically about five of each per session), so we do not expect them to differ simply because of drifts in behavior or neural activity over time (Chaisanguanthum et al., 2014). Still, the context of these blocks could be sufficient to change neural activity or behavior (Verstynen and Sabes, 2011). Therefore, we compared the firing rates for targets that appeared in both trial block types, for both the delay period and the peri-movement period (Fig. 2). The firing rates are highly correlated across blocks, with correlation coefficients of >0.95 , indicating that the neural activity patterns for a given reach are independent of block type.

Results

We recorded from large numbers of neurons in PMd and M1 cortices as two monkeys performed a “center-out” reaching task, in which targets appeared at one of eight locations arrayed 7 cm from a fixed start point. Our goal here is to study the role that these cortical areas play in transforming sensory signals into the appropriate actions. To simplify the problem, we focus on the relationship between preparatory neural activity, recorded during an instructed delay period, and the initial movement trajectory (see Materials and Methods). Although we will ultimately build a causal model of the transformation from perception to action, we begin with a more traditional “tuning curve” analysis that will help motivate the causal model.

Distinguishing target and movement tuning

Many studies have used center-out reaching to fit tuning curves that relate PMd/M1 firing rates to either the target angle or the initial movement angle (Georgopoulos et al., 1982; Schwartz et al., 1988; Schwartz and Moran, 2000). However, there has been little analysis of how these two relate. Consider target tuning first. Figure 3, *a* and *b*, shows standard cosine tuning curves as a func-

tion of target angle (Eq. 1) for two example neurons in our dataset, fit using data from the eight-target trial blocks (see Materials and Methods). These tuning curves can be thought of as the first-order (i.e., linear) approximation of the mapping from target location to preparatory activity (Sanger, 1996). Because cosine tuning can be used to predict movement velocity, this target tuning function has been interpreted as a readout of the intended movement direction or the causal movement command signal (Georgopoulos et al., 1982; Schwartz, 2004; Georgopoulos and Carpenter, 2015). Similar arguments have been made for cosine tuning in local field potentials and surface electrocorticography in motor cortex (Mehring et al., 2003, 2004). However, the standard center-out reaching paradigm is a poor tool for distinguishing between target tuning and movement tuning, since the target and movement directions covary so strongly.

By having our animals perform large numbers of repeated reaches to each target in the three-target blocks (see Materials and Methods), we can study the relationship between preparatory firing rates and movement direction independent of target location. Specifically, we performed a perturbative analysis, where for each cell and each target we regress variation in the firing rate against variation in the initial movement angle (Eq. 3). The linear fit represents a first-order approximation of the true relationship, for a given single target, between firing rate and movement variability. We focus on the “movement tuning slope” f_i for each target and each cell (Eq. 3). Movement tuning fits are shown for the same example neurons in Figure 3, *c* and *d*.

If target and movement tuning are indeed just different measures of the same mapping, then we would expect that for each cell and each target, the movement tuning slopes f_i would match the slopes of the cosine target tuning function. For the cell in Figure 3*c*, the target and movement tunings slopes do match well, meaning that the neuron’s activity covaries with the target location angle and the initial movement angle in the same way. However, this relationship does not always hold. In Figure 3*d*, the target and movement tuning slopes appear to be out of phase with each other, implying the cell’s activity covaries oppositely with target angle and initial movement angle.

Figure 3, *e* and *f*, shows the relationship, across the population of recorded neurons and across targets, between target tuning slopes and movement tuning slopes. Perhaps counterintuitively, the correlation between the two slopes appears to have near-zero

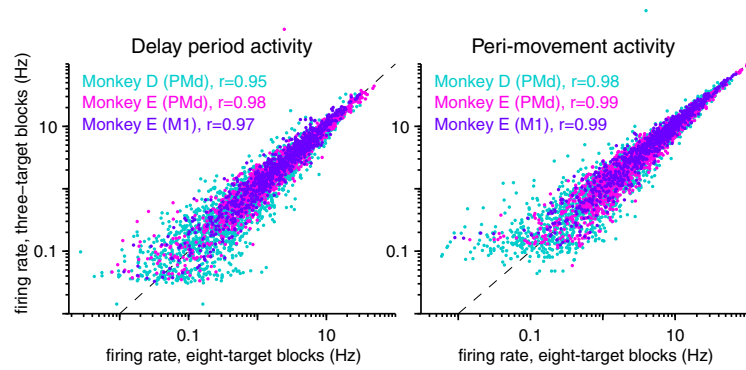


Figure 2. Neural activity does not exhibit context dependence between experimental block types. Scatter plots show a comparison of the mean firing rates in the three-target and eight-target blocks. Each data point is the mean rate for a given recorded cell and target, during the delay period (left) or the peri-movement period (right). Log–log plots are shown for clarity, but the Pearson correlation coefficients were computed on the raw values. Similar results ($p \geq 0.94$) were obtained with Spearman rank-order correlations.

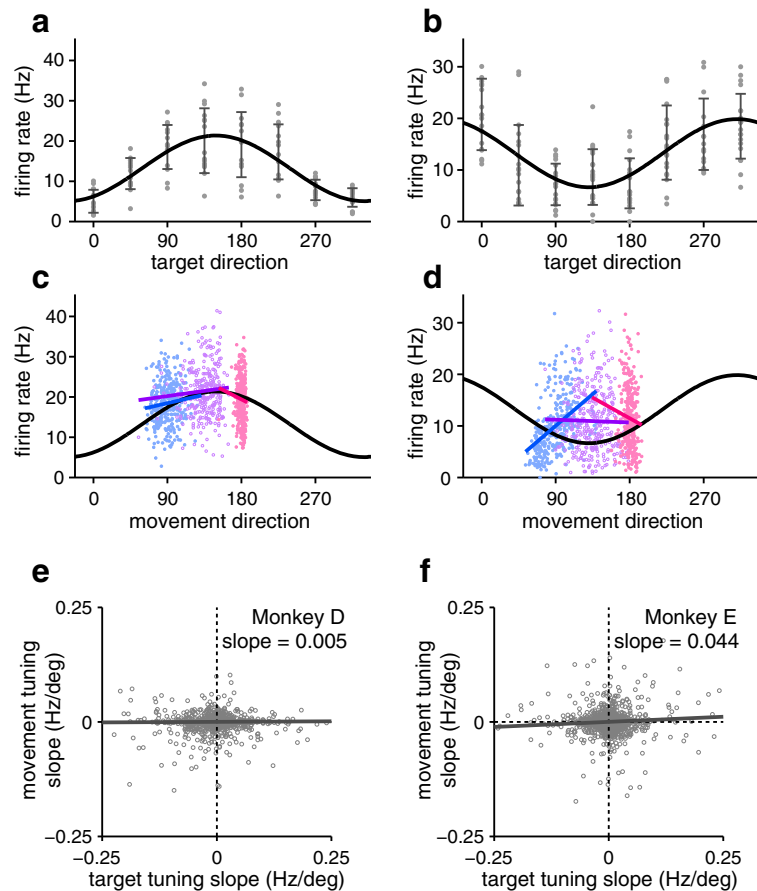


Figure 3. Relationship between target and movement tuning. *a, b*, Target tuning. For two example cells (Monkey E, same experimental session), delay-period firing rate is plotted against the reach target angle (gray points represent trials; open circles and error bars show per-target means and SDs, respectively) for trials in the eight-target blocks. The black trace is the best-fit cosine tuning curve (Eqs. 1, 4). *c, d*, Movement tuning. Scatter plots show delay-period firing rates versus initial movement angle for individual trials, color-coded by target. For each target in the three-target blocks, the reach regression slope (Eq. 3) is drawn. The cosine target tuning curve from *a* and *b* is shown for comparison in black. *e, f*, Scatter plot of reach slope versus target slope for each cell and each target in the three-target blocks of all sessions.

correlation: the way an individual neuron’s activity covaries with target angle has little bearing on how it covaries with initial movement. In the following, we try to make sense of this result in terms of a more explicit casual model for how PMd/M1 implements the mapping from sensory input to motor output.

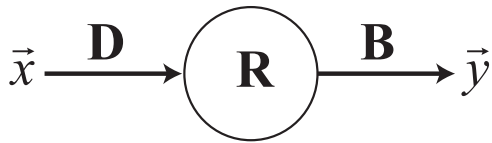


Figure 4. A simple schematic of the circuit that drives visually guided reach. A stimulus \vec{x} is presented; information about this stimulus is encoded by the firing rates R of cells in a neural population, which in turn drive a motor command \vec{y} . \mathbf{D} is the target tuning matrix (Eq. 4), and \mathbf{B} is the motor weighting matrix (Eq. 5).

Neural tuning curves in the context of a causal model of movement initiation

A linear causal model

Here, we model the causal link between perception and action in two steps: how sensory cues such as the target angle drive preparatory activity, and how that activity affects movement variables such as the initial movement angle (Fig. 4). In both cases, we use linear approximations to the true mapping, since we will apply these only to variations about the mean (firing rate and initial movement) for a given target.

The first causal step is how sensory inputs, in this case the two-dimensional target location \vec{x} , drive the firing rates \mathbf{R} of a population of n cells:

$$\mathbf{R} = \mathbf{D}\vec{x} + \varepsilon, \quad \varepsilon = N(0, \mathbf{E}), \quad (4)$$

where trial-to-trial variability in the firing rates, ε , is modeled as a normal distribution with covariance matrix \mathbf{E} and each row of \mathbf{D} (an $n \times 2$ matrix) is a vector representing a given cell's target tuning, with the angle of the vector representing the preferred direction and length determining the modulation depth and speed sensitivity. Note that for center-out reaching, Equation 4 is simply the matrix notation for a population of cells with cosine tuning in target angle (Eq. 1), except that variability can be correlated across neurons. This term will model variability that is likely attributable to multiple sources, including variations in the afferent signals to PMd/M1 as well as local neural "noise" (e.g., Poisson sampling), with correlations between cells arising from common neural inputs as well as local circuit dynamics. Whereas estimation of the full covariance structure of \mathbf{E} may be data limited, it is straightforward to estimate \mathbf{D} and the diagonal elements of \mathbf{E} via linear regression on a per-cell basis.

The second step is to model the causal link from PMd/M1 preparatory activity \mathbf{R} to the initial velocity vector \vec{y} . We avoid the very difficult prospect of modeling both local circuit dynamics and downstream neuromuscular dynamics, with a perturbative model that approximates the relationship between variability in neural activity and variability in movement for a single target \vec{x}_0 :

$$\vec{y} - \vec{y}_0 = \mathbf{B}^T(\mathbf{R} - \mathbf{R}_0) + N(0, \Sigma_{\text{downstr}}), \quad (5)$$

where \mathbf{R}_0 and \vec{y}_0 are the mean firing rates and initial velocity, respectively, for reaches to \vec{x}_0 and each row of \mathbf{B} (an $n \times 2$ matrix) is a vector representing the effect a cell's preparatory activity has on y , i.e., its weighting in the motor output for reaches to this target. Here, the covariance matrix Σ_{downstr} models all of the movement variability not ascribable to neuronal variability, which can arise downstream, e.g., through motor noise (Harris and Wolpert, 1998; Todorov and Jordan, 2002), or otherwise independently of the PMd/M1 population that we record. The motor weighting matrix \mathbf{B} here is not simply a statistical quantity but is intended to reflect the underlying biophysical system; the rows in \mathbf{B} determine how the motor output would change from

one trial to the next if the response of only a single neuron were changed.

Unlike the target tuning matrix \mathbf{D} , the motor weighting matrix \mathbf{B} cannot be easily measured from the data. This is because the problem is poorly conditioned: the number of neurons (n) is large compared with the low dimensionality of the behavioral measure (2): simply requiring $\mathbf{B}^T\mathbf{D} = \mathbf{I}$, i.e., the movement to match the cue, leaves the inference of \mathbf{B} severely underconstrained. For example, adding or subtracting a small subset of neurons would have a large impact on the best-fit b_i for the remaining neurons. Of course, one can find a linear population "decoder" \mathbf{Q} , such that $\vec{y} = \mathbf{Q}^T(\mathbf{R} - \mathbf{R}_0)$ provides an empirical mapping from neural population activity to the predicted initial movement vector \vec{y} . But because this matrix is underconstrained, good movement prediction does not imply that \mathbf{Q} is a good predictor of \mathbf{B} . Furthermore, because we are, in practice, unable to record from every relevant neuron, the entries in \mathbf{Q} also include the effects of all missing neurons that are correlated with those in our population.

The fact that \mathbf{B} cannot be directly measured or inferred from the data motivates the rest of this work.

Movement tuning and its relationship to the causal model

Instead of estimating the motor weighting matrix \mathbf{B} from the data, we will estimate the inverse mapping, i.e., the movement tuning model that describes firing rates as a function of initial velocity. Here again, we model perturbations around the mean for repeated reaches to the same target, \vec{x}_0 :

$$\mathbf{R} - \mathbf{R}_0 = \mathbf{F}^T(\vec{y} - \vec{y}_0) + \text{noise}, \quad (6)$$

where each row of \mathbf{F} represents the movement tuning for a given cell. (Note that movement tuning is fit independently for each cell and so is different from the putative decoder matrix \mathbf{Q} described above.) For our center-out reaching task, Equation 6 simplifies to the cosine tuning for reach angle, as in Equation 3 (see Materials and Methods) and Figure 3. For each recording session, target, and cell, we fit this model and, in particular, the movement tuning slope f_i , using simple linear regression. Examples were shown in the linear fits of Figure 3, *c* and *d*.

Whereas Equation 6 is intended to capture correlation, not causality, the causal model of Equations 4 and 5 provides an explicit link between the movement tuning matrix \mathbf{F} and the motor weighting matrix \mathbf{B} : with Equation 5, we can write the solution to the regression problem of Equation 6 as follows:

$$\mathbf{F} = \langle \varepsilon(\vec{y} - \vec{y}_0)^T \rangle \langle (\vec{y} - \vec{y}_0)(\vec{y} - \vec{y}_0)^T \rangle^{-1} = \mathbf{E}\mathbf{B}\Sigma_{\text{tot}}^{-1}, \quad (7)$$

where angle brackets denote averages over repeated reaches to target \vec{x}_0 and $\Sigma_{\text{tot}} = \mathbf{B}^T\mathbf{E}\mathbf{B} + \Sigma_{\text{downstr}}$ is the 2×2 total movement covariance matrix, which is the sum of the effects of neuronal variability and downstream movement variability and which can be measured directly from data.

We cannot use Equation 7 to infer the causal model parameters from cells' movement tuning, i.e., to infer \mathbf{B} directly from \mathbf{F} , as this would require knowledge of the full neuronal covariance matrix, \mathbf{E} , including all neurons that affect behavior, which is, of course, experimentally intractable. Instead, we predict values for \mathbf{B} , and hence for \mathbf{F} , using an experimentally supported normative model.

Neural tuning curves prediction with a normative model

A minimum-variance causal model

Since we cannot directly estimate the motor weights \mathbf{B} of Equation 5 from the fit tuning curves \mathbf{F} , we instead ask what values of

B would be best, given the empirically measured target tuning parameters **D**. To accomplish this, we posit the normative principle of minimal variance in sensori-motor control, which has considerable empirical backing (Harris and Wolpert, 1998; Todorov and Jordan, 2002; Todorov, 2004; van Beers, 2009). Specifically, we hypothesize that the map **B** between population activity and behavior should generate maximally precise movements, while also yielding an unbiased estimate, i.e., $\mathbf{B}^T \mathbf{D} = \mathbf{I}$. It is easily shown, by the method of Lagrange multipliers, that the solution for the map **B** that minimizes the trace of the covariance matrix Σ_{tot} in the model defined by Equations 4 and 5 is as follows:

$$\mathbf{B} = \mathbf{E}^{-1} \mathbf{D} [\mathbf{D}^T \mathbf{E}^{-1} \mathbf{D}]^{-1}. \quad (8)$$

(This form minimizes any linear function of Σ_{tot} as well as its determinant.) The (minimized) contribution of the neural population to the behavioral covariance is thus $\Sigma_{\text{pop}} = \mathbf{B}^T \mathbf{E} \mathbf{B} = [\mathbf{D}^T \mathbf{E}^{-1} \mathbf{D}]^{-1}$. To gain intuition for Equation 8, note that were **E** isotropic (independent and identically distributed neural variability), the optimal motor weighting matrix **B** would be the pseudo-inverse of the target tuning matrix **D**. If **E** were instead only diagonal (independent neural variability), then from the numerator of Equation 8 we see that the motor weighting vectors \tilde{B}_i would be scaled inversely by the variance of each neuron's preparatory firing rates, akin to the well known solution for minimum-variance cue combination (Ernst and Banks, 2002).

The minimum-variance principle (Eq. 8), when combined with Equation 7, leads to the primary prediction of this model:

$$\mathbf{F} = \mathbf{D} \Sigma_{\text{pop}} \Sigma_{\text{tot}}^{-1} = \mathbf{D} (\mathbf{I} + \Sigma_{\text{downstr}} \Sigma_{\text{pop}}^{-1})^{-1}, \quad (9)$$

where **I** is the identity matrix. In words, the minimum-variance principle predicts that the target tuning of each cell is related to its movement tuning by a simple two-dimensional, positive definite linear transformation, $\Sigma_{\text{pop}} \Sigma_{\text{tot}}^{-1}$, which is constant across the population (though it may be different for each target).

Predictions of the minimum-variance model

To get an intuition for Equation 9, consider the limiting case where there is no downstream noise. In this case, the target and movement tunings should be identical, up to fitting noise. More generally, the simple linear relationship mapping **D** to **F** implies that the slope of neurons' movement tuning, f_i , should correlate (across cells and targets) with the local slopes of their target tuning curves d_i .

The idea that a neuron should be similarly tuned to the cued and actual movement is consistent with the intuition of population coding: if a task cue drives activity in a subset of neurons, those neurons should drive the movement in the appropriate directions. Here, it might seem that the minimum-variance principle provides a normative justification of this intuition. However, as described above, this prediction does not appear to be consistent with our data: Figure 3, *e* and *f*, shows a very weak correlation, if any, between the target and movement tuning slopes across our large population of cells.

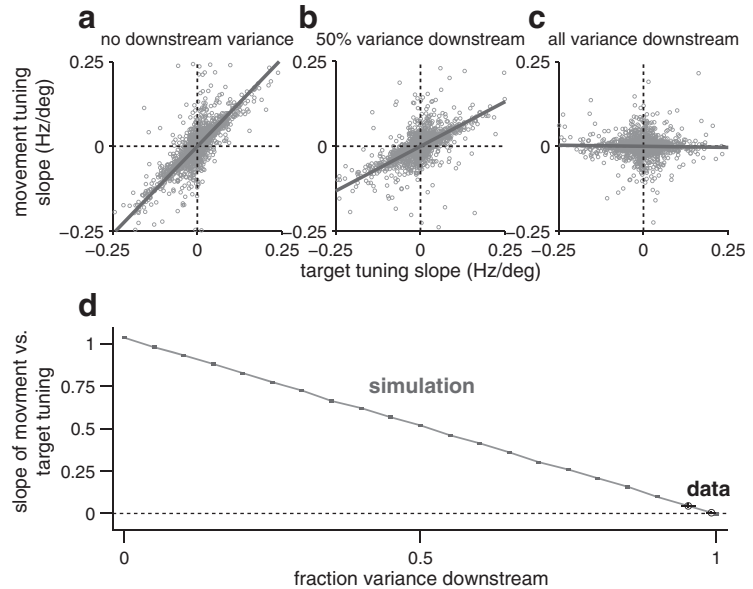


Figure 5. Simulation of the effect of downstream noise. **a–c**, The expected relationship between target and reach slopes predicted by the minimum-variance model (Eq. 9) for 0% (**a**), 50% (**b**), and 100% (**c**) of total motor variance arising downstream of the neural population. Here, behavioral noise caused by population noise ($\Sigma_{\text{pop}} = [\mathbf{D}^T \mathbf{E}^{-1} \mathbf{D}]^{-1}$) and downstream contribution (Σ_{downstr}) are assumed to be isotropic. **d**, Summary: expected regression slope between target and reach slopes (SE shown) as a function of fraction of motor variance that arises downstream of the neural population. Black circles (with errors) show where experimental data (see Fig. 3) lie.

One other notable feature of Equation 9 is the different (and unrelated) origins of scale between the elements of the tuning coefficients **F** and the motor weighting matrix **B**. Given the large population of relevant motor cortical neurons, the rows of **B** should scale with **D** as $1/n$, on average. In contrast, the magnitude of the movement tuning coefficients **F** are related to those of the target tuning parameters **D** by a transformation that depends only on the distribution of sources of variability. If κ is some measure of the fraction of motor variability manifest in this neural population (i.e., not downstream), then $\Sigma_{\text{pop}} \Sigma_{\text{tot}}^{-1} \sim \kappa$. Thus, the coefficients in **F** will be of order κ times the size of the elements in **D**. Given the large number of motor cortical neurons, we presume that $\kappa \gg 1/n$.

If downstream variability is the dominant contributor to the total movement variability, then $\Sigma_{\text{pop}} \Sigma_{\text{tot}}^{-1}$ will be small, and so the slope of neurons' movement tuning, f_i , will be small and the correlation between f_i and d_i is expected to be weak. This possibility is investigated next.

Effects of downstream noise

Here, we consider whether the presence of downstream variability could explain the lack of correlation between movement and target tuning slopes. We address the problem by simulating datasets with different downstream noise levels. Examples are shown in Figure 5*a–c*. As expected, both the strength of correlation and the regression slope between d_i and f_i decrease as the fraction of downstream noise increases. This relationship is summarized in Figure 5*d*, which shows a linear decrease in the d - f slope as a function of the fraction of movement variability attributable to downstream noise.

We can use the linear relationship of Figure 5*d* to estimate the contribution of downstream noise, given the empirical d - f slope from Figure 3, *e* and *f*. To explain the nearly absent correlation between d_i and f_i observed in the empirical data, nearly all of the behavioral variability (99% for Monkey D, 95% for Monkey E)

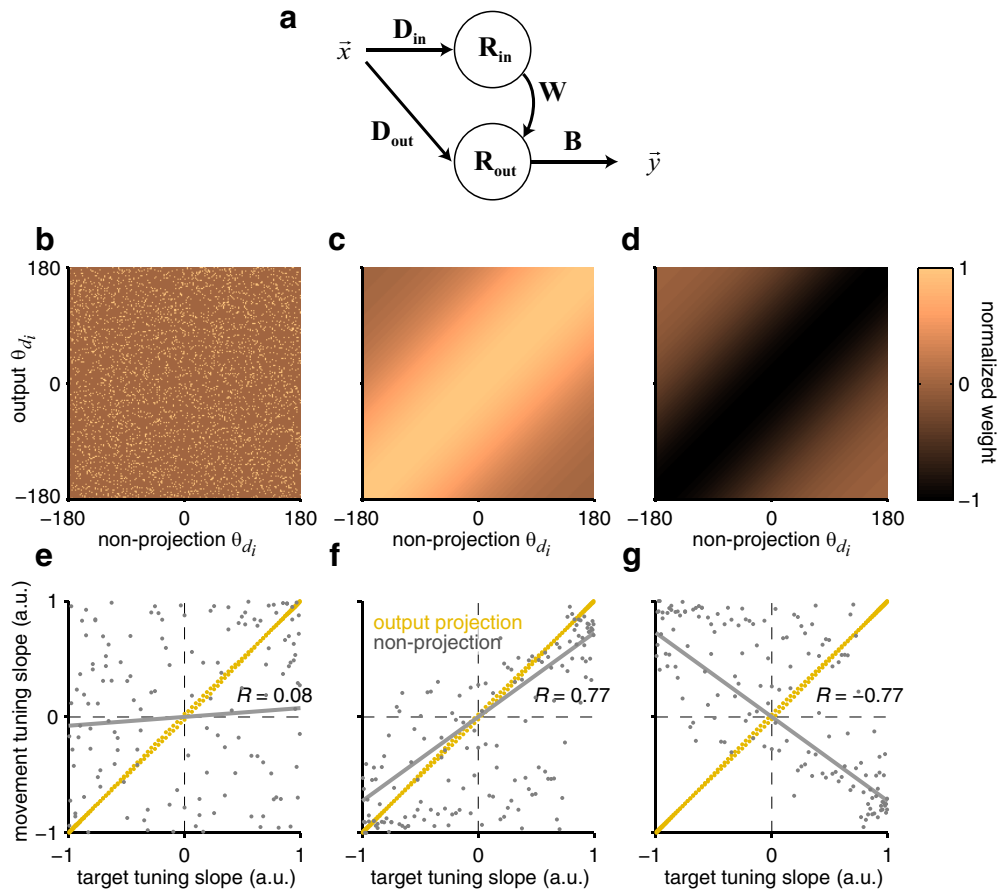


Figure 6. Effect of nonprojection neurons. **a**, An augmented model for the sensorimotor circuit. As in Figure 4, behavior is driven by neural activity encoding the stimulus \bar{x} ; however, here there is also a subpopulation of input cells or interneurons whose activity R_{in} only affects the output of the circuit through connections, via weight matrix W , to the population of output projection neurons. **b–d**, Example weight matrices W for the augmented model in **a**: random connectivity between nonprojection neurons and output neurons (**b**), tuned connectivity with positive weights (**c**), and tuned connectivity with negative weights (**d**). (For display, weights are normalized to range between positive and negative unity.) **e–g**, Simulated minimum-variance predictions for the relationship between target and reach slopes for the output projection (gold; Eq. 12a) and nonprojection (gray, with regression slope; Eq. 12b) cell populations. These simulations do not include downstream or measurement/statistical noise.

would have to arise downstream of PMd/M1 (empirical data points in Fig. 5d). This estimate is not consistent with previous studies showing significant cortical contributions to motor variability (Churchland et al., 2006; Huang and Lisberger, 2009). Some of the discrepancy may be caused by movement variability not captured by the model of Equations 4 and 5, including error attributable to the linear approximations in the model. However, by focusing on variation about the mean initial movement for repeated trial conditions, our experimental design minimizes such error, and so we do not expect that this could explain the discrepancy. Rather, the result suggests that either the minimum-variance hypothesis or the simple causal model of Figure 4 is wrong. We next show how a simple addition to the causal model can rescue the minimum-variance hypothesis and provide new insight into the causal link from PMd/M1 activity to motor output.

Causal models with nonprojection neurons

The simple, feedforward model of Equations 4 and 5 and Figure 4 assumes that the whole population of neurons we recorded in PMd/M1 are input sensory-receiving neurons as well as motor-projecting neurons. In reality, the intrinsic circuitry of PMd and M1 has substantial structure, with task-related sensory inputs (Wise et al., 1983; Cisek et al., 2003; Song and McPeck, 2010), cortical output units originating in layer 5, as well as extensive

connectivity within PMd/M1 mediated by both excitatory cells and inhibitory interneurons (Jones and Wise, 1977; Huntley and Jones, 1991; Tokuno and Nambu, 2000). Figure 6a shows an incrementally more realistic schematic that incorporates some of this structure, while remaining amenable to simple analysis. This schematic has two subpopulations of neurons: output projection neurons that have direct effects on movement with rate vector R_{out} and nonprojection neurons with rate vector R_{in} , where the subscript “in” reflects the fact that these are sensory input neurons, interneurons, or other intracortical neurons. These nonprojection neurons only affect movement through their connections to the projection neurons.

We formalize the schematic of Figure 6a with a simple extension to the causal model of Equations 4 and 5 that includes these two separate populations:

$$R_{in} = D_{in}^T \bar{x} + N(0, E_{in}), \quad (10a)$$

$$R_{out} = D_{out}^T \bar{x} + W R_{in} + N(0, E_{out}), \quad (10b)$$

where D_{in} and E_{in} are the nonprojection target tuning matrix and firing rate covariance, respectively, and the matrix W describes the connections from the nonprojection neurons to the output projection neurons. D_{out} and E_{out} in Equation 10b are the target tuning and covariance matrices that would be observed in pro-

jection cells if the nonprojection inputs were silenced. (Note that \mathbf{D}_{out} would be 0 if neurons that receive sensory input are a distinct population from those that directly effect downstream movement, since the causal effect of \tilde{x} on R_{out} is only through the activity of the input neurons). Combining Equations 10a and 10b yields the effective target tuning and covariance for the output projection neurons in the intact network:

$$\mathbf{R}_{\text{out}} = \mathbf{D}_{\text{eff}}^T \tilde{x} + N(0, \mathbf{E}_{\text{eff}}). \quad (11)$$

Equation 11 shows that the causal model for the output projection neurons alone has the same structure as the original model of Equations 4 and 5, with the target tuning and covariance matrices replaced by the effective parameters: $\mathbf{D}_{\text{eff}} = \mathbf{D}_{\text{out}} + \mathbf{D}_{\text{in}} \mathbf{W}^T$ and $\mathbf{E}_{\text{eff}} = \mathbf{E}_{\text{out}} + \mathbf{W} \mathbf{E}_{\text{in}} \mathbf{W}^T$. It is these effective matrices that would be estimated empirically from the target tuning regression.

We can next use Equations 10 and 11 to predict the relationship between target and movement tuning. For output neurons, the prediction is analogous to that derived from the original model: if we define $\Sigma_{\text{pop,eff}} = (\mathbf{D}_{\text{eff}}^T \mathbf{E}_{\text{eff}}^{-1} \mathbf{D}_{\text{eff}})$, then the movement tuning matrix should be as follows:

$$\mathbf{F} = \mathbf{D}_{\text{eff}} \Sigma_{\text{pop,eff}}^{-1} \Sigma_{\text{tot}}^{-1}. \quad (12a)$$

For the nonprojection neurons, the result takes on a different form, since variability in these cells only affects behavior through the output neurons:

$$\mathbf{F}_{\text{in}} = \mathbf{E}_{\text{in}} \mathbf{W}^T \mathbf{E}_{\text{eff}} \mathbf{D}_{\text{eff}} \Sigma_{\text{tot}}^{-1}. \quad (12b)$$

Notably, the movement tuning matrix for nonprojection cells, \mathbf{F}_{in} , depends on the nonprojection target tuning matrix \mathbf{D}_{in} only through the effective target tuning of the output neurons, \mathbf{D}_{eff} .

Figure 6*b–d* illustrates the predictions of Equations 12a and 12b for three intuitive examples of the projection matrix \mathbf{W} . For each case, we simulated a network with 200 output cells and 200 nonprojection cells, with target tuning vectors evenly spaced around the circle, with fixed modulation depth, and with pairwise neural correlations randomly drawn from the distribution of correlations observed in data. Since the goal of the simulation is to explore the effects of the projection matrix on the relationship between target and movement tunings, downstream noise (which would simply weaken this relationship) was not included.

As a first example (Fig. 6*b*), we considered projection matrices \mathbf{W} where nonprojection neurons are randomly (and sparsely) connected to output neurons, in a way that is independent of their target tunings. This connectivity randomizes the effect of the nonprojection neurons on the output, so that there is little observed relationship between the target and movement tuning for this population (Fig. 6*e*, gray data points). In contrast, there is still a tight correlation between target and movement tuning for the output population (Fig. 6*e*, yellow data points), as expected for the case of no output noise (compare Fig. 5*a*).

Second, we considered the case where nonprojection neurons connect to output neurons with similar target tuning angles. Specifically, the elements of \mathbf{W} were chosen to be Gaussian in the difference between their preferred target directions (Fig. 6*c*). In this case, we found that the target and movement tuning are correlated for nonprojection cells, albeit not as strongly as for output cells (Fig. 6*f*). From the perspective of nonprojection neurons, the diffuse connectivity with and neural variability in the output neurons acts as “downstream noise” (compare Fig. 5*b*).

Finally, we considered the case where nonprojection neurons connect to output neurons with similar target tuning angles, as above, but with negative weights, as if the nonprojection neur-

ons were primarily inhibitory interneurons (Fig. 6*d*). In this case, there is negative correlation between the target and movement tuning slopes (Fig. 6*g*).

We stress that the linear models in Figure 6 are not intended to be taken literally, as accurate descriptions of the circuit connectivity. Rather, they are meant as schematic descriptions of the dominant feedforward cortical connectivity patterns that might underlie sensorimotor transformations in PMd/M1. These patterns, along with the first-order approximation of their activity, lead to different predicted relationships between target and movement tuning slopes. The results of this analysis (Fig. 6*e–g*) suggest that the lack of correlation we observe between these slopes (Fig. 3) may be because our dataset includes a heterogeneous sampling of cells. We consider this possibility next.

Comparison of target and movement tuning in neural subpopulations

The results in Figure 6 suggest that target and movement tuning correlations should be computed separately for functionally distinct cell types, such as input neurons, inhibitory interneurons, and output projection neurons. Of course, such information is not directly available from our extracellular recordings. Some reports have indicated that cell types can be identified from the shape of the action potential waveform recorded extracellularly in macaque, specifically from the trough-to-peak time (Mountcastle et al., 1969; Mitchell et al., 2007; Merchant et al., 2008; Cohen et al., 2009; Kaufman et al., 2010; Song and McPeck, 2010; Yokoi and Komatsu, 2010). However, more recent evidence suggests that in PMd and M1, whereas different cell types have different waveform shapes in the aggregate, waveform shape does not provide reliable identification of cells at the level of individual neurons (Vigneswaran et al., 2011). Therefore, rather than attempting to positively categorize cells, we bin them coarsely by trough-to-peak time (Fig. 1), with the expectation that different bins may contain very different distributions of cell types.

We examine the relationship between target and movement tuning separately for each waveform bin. One might expect that in the short waveform bin ($<300 \mu\text{s}$ trough-to-peak), a high proportion of interneurons (Mountcastle et al., 1969; Mitchell et al., 2007; Merchant et al., 2008; Kaufman et al., 2010) would yield a negative correlation between target and movement tuning, as in the synthetic example of Figure 6*g*. Conversely, for longer waveform bins, one might expect that a large proportion of pyramidal cells, including output pyramidal tract neurons, would yield positive correlations between target and movement tuning, as for the output neurons in Figure 6*e–g*.

The results for three representative waveform bins are shown in Figure 7*a–c*. For both short and long waveforms (200–250 and 500–550 μs peak-to-trough; Fig. 7*a,c*), there is no correlation between target and moving tuning slopes. However, for cells with intermediate waveforms (350–400 μs peak-to-trough; Fig. 7*b*), there is a significant and positive correlation between the two tuning slopes. Summary data are shown in Figure 7, *d* and *e*. For each animal and each brain area, there is no significant correlation between target and movement tuning for short ($<300 \mu\text{s}$) or long ($>400 \mu\text{s}$) waveforms. (One exception is the shortest bin, 150–200 μs for PMd in Monkey E, where there was a small negative correlation.) In contrast, for both monkeys and both brain areas, we observe a peak in the slope relating movement and target tuning for intermediate waveforms, with significantly positive correlations between tunings in the 350–400 μs bin. This bin contains 3, 11, and 12% of recorded cell sessions in Monkey D PMd, Monkey E PMd, and Monkey E M1, respectively (Fig. 1).

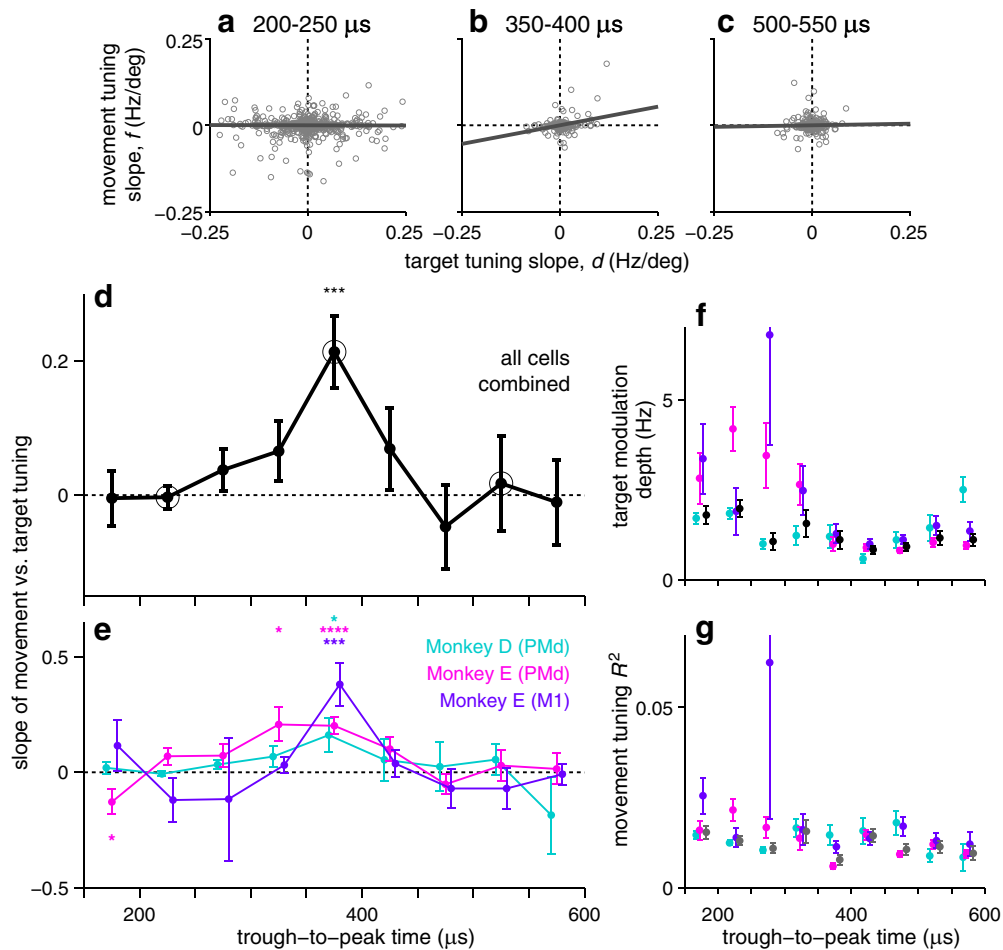


Figure 7. Differences in target and movement tuning correlation across spike waveform shape. **a–c**, Scatter plot of movement tuning slope f versus target tuning slope d for cells in each of three waveform groups, defined as trough-to-peak times from 200–250 μ s (**a**), 350–400 μ s (**b**), and 500–550 μ s (**c**). Data are for each target in the three-target blocks of each session and for both animals combined (compare Fig. 3); gray lines are regression fits to the f versus d data. **d**, Regression slope of f versus d for subsets of cells identified by their waveform trough-to-peak time, in 50 μ s bins; data from both animals are combined. Open circles denote the datasets shown in **a–c**. **e**, Same as **d**, but with data separated by animal and brain area. Asterisks denote statistical significance ($*p < 0.05$, z test) of nonzero result; additional asterisks denote additional SE of significance, i.e., $***p < 10^{-4}$ and $****p < 10^{-6}$. **f, g**, For cells binned as in **d** and **e**: mean (and SEM) of target tuning modulation depth ($|\bar{d}_i|$ in Eq. 1; **f**) and coefficient of determination R^2 of the movement tuning regression (**g**). Note that for **g** only, the regression was computed on the first-order difference (trial-to-trial changes) in both neural activity and behavior, as this mitigates potential effects of nonstationarity (Chaisanguanthum et al., 2014), providing a more unbiased estimate of the strength of tuning. Results are qualitatively unchanged when the raw values are used instead.

Note that because we high-pass filter our electrophysiological signals, the reported trough-to-peak times are somewhat faster than the true physiological values (Gold et al., 2006; Quian Quiroga, 2009; Vigneswaran et al., 2011). We estimate the true trough-to-peak times of the waveforms in the 350–400 μ s bin to be \sim 460–520 μ s (Gold et al., 2006; Quian Quiroga, 2009).

Since only one bin in Figure 7*d* shows a significantly positive slope, it is important to demonstrate that the effect is not spurious. First note that we obtain highly consistent results if we consider the regression coefficient of determination (R^2) instead of regression slopes. Moreover, the computed significance level of the 350–400 μ s bin in Figure 7*d* is $p < 0.0001$, which with a naive multiple-comparisons correction remains at the $p < 0.001$ level. Furthermore, the removal of the three apparent outlying points in first quadrant of Figure 7*b* does not affect the significance of the positive slope. More importantly, that same bin is significant for each of three recording arrays (Fig. 7*e*). Indeed, the probability of getting three spurious false positives in the same time bin, at only the $p < 0.05$ level, is $p = 0.001$.

We also need to ensure that the results of Figure 7, *d* and *e*, are not simply artifacts of the degree of target or motor tuning

across cells groups. Therefore, we also examined the target tuning strength, quantified as the modulation depth $|\bar{D}_i|$ in Figure 7*f*, and the movement tuning strength, quantified as the coefficient of determination R^2 for neuron-behavior correlations in Figure 7*g*. There is no pattern in the modulation depths or neuron-behavior R^2 values in Figure 7, *f* and *g*, that would appear to explain the slopes in Figure 7, *d* and *e*. In fact, the smaller R^2 values in the 350–400 μ s bin in Figure 7*g* might lead one to conclude that these cells are the least directly connected to the motor output, the typical interpretation of neuron-behavior correlations. However, these R^2 values conflate the empirical variability of a cell and its direct effect on behavior. Our modeling work shows that this confound can be circumvented: cells that contribute more directly to the motor output can be identified through the relationship between target and movement tunings.

Intermediate waveform durations and motor output

By the logic of the toy models in Figure 6, the positive correlations between movement and target tuning for intermediate waveforms suggests that these bins have a large concentration of output pyramidal neurons, or at least that these cells have the most

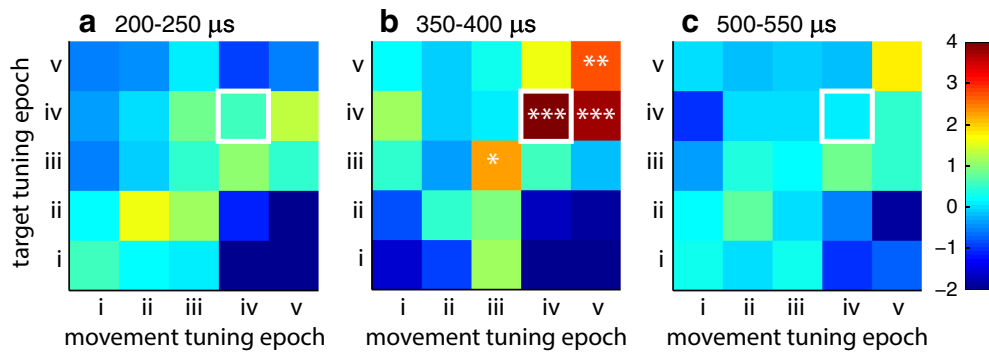


Figure 8. Relationship between movement tuning and target tuning, with tunings measured across and between epochs. Each square represents the normalized slope (i.e., divided by the SE) of the regression between a target tuning d and a movement tuning f , where d and f come from the epoch noted on the ordinate and abscissa, respectively. The epochs are numbered as follows: (i) pretrial, (ii) trial start, (iii) trial ready, (iv) delay period (premovement), and (v) peri-movement (see Materials and Methods). **a–c** were computed with the same subsets of cells used in Figure 7*a–c*. Thus, the three squares outlined in white correspond to delay-period data marked with open circles in Figure 7*d*. Squares marked with asterisks correspond to those with significantly positive slopes ($*p < 0.02$, $**p < 0.005$, $***p < 0.0001$).

direct effect on behavior. This interpretation is supported by the observation that for Monkey E, where we recorded from both M1 and PMd, the slope relating movement and target tuning is significantly larger for M1 than PMd ($p < 0.05$, z test).

Assuming that the 350–400 μ s bin contains mostly output neurons, we can compare the empirical slopes for that bin in Figure 7*d* to the simulated values in Figure 6*d* to estimate the relative contributions of neural and downstream variability to total movement variability. With this approach, we estimate that variability in PMd (and upstream) contributes 16 ± 7 and $20 \pm 4\%$ to the overall movement variability in Monkeys D and E, respectively, and variability in M1 (and upstream) contributes $38 \pm 10\%$ for Monkey E. These numbers are consistent with previously reported values measured in different ways: van Beers (2009) used motor learning in a human reach task, Churchland et al. (2006) directly measured neuron-behavior correlations (in a different type of task), and in another study, we estimated this quantity from autocorrelations in behavior (Chaisanguanthum et al., 2014).

If this intermediate waveform bin contains many output cells, then their causal effect should persist into the movement period. Therefore, we predict a correlation between target movement tuning for these cells during the peri-movement epoch as well. In contrast, no such correlation should be observed before movement planning begins. To test these predictions, we repeated the analysis of Figure 7*a–d* using data from six different trial epochs (see Materials and Methods). Results are shown in Figure 8. For the subpopulation of cells with intermediate waveform durations (350–400 μ s; Fig. 8*b*), we find that there is significant correlation between target and movement tuning in the last three trial epochs. We also find that the target tuning during the delay period correlates strongly with the movement tuning during the peri-movement period, but not vice versa, consistent with a causal flow of information. In contrast, no correlations between target and movement tuning are observed early in trial activity nor for the shorter or longer waveform cell populations (Figs. 8*a,c*).

As noted above, fast waveforms have previously been associated with motor cortical interneurons (Mountcastle et al., 1969; Mitchell et al., 2007; Merchant et al., 2008; Kaufman et al., 2010). Here, however, we only see a significant negative correlation between movement and target tuning for the shortest waveform bin and in only one dataset (Monkey E PMd). This suggests that either that the interneurons do not have a strong and tuned influence on output neurons or that the short waveform bins are not dominated by in-

terneurons, as seen in the study by Vigneswaran et al. (2011). Neurons with the longest waveforms might represent pyramidal cells with smaller somata (Vigneswaran et al., 2011) and, therefore, may not include a predominance of output neurons. The lack of correlation between movement and target tuning would be expected if, for example, the role of these cells were to coordinate the intrinsic dynamics of motor cortex (Churchland et al., 2010), rather than to directly drive movement.

Discussion

Tuning curves in the context of causal models

We have analyzed the preparatory neural activity of cells recorded in PMd and M1 and have found that target and movement tuning are not correlated across cells and targets. This observation is at odds with an implicit assumption in the field, dating to at least Georgeopoulos et al. (1986), that spatial tuning is a singular property of cortical neurons. In contrast, recent models have emphasized the causal role that PMd/M1 activity plays in movement generation, de-emphasizing the sensory tuning of these cells (Todorov, 2000; Churchland et al., 2010; Lillicrap and Scott 2013; Shenoy et al., 2013). Here, we have attempted to link these approaches, using a simple, linear causal model.

Whereas a general first-order model is underconstrained, allowing for accurate movement with many combinations of target and movement tuning, concrete predictions can be made by adopting the commonly used and empirically supported principle of minimum-variance movement control. With this model, we find that the movement and target tuning are only expected to correlate for output cells that directly project to the motor periphery. Therefore, we separately analyzed subpopulations of neurons, defined by their spike waveform lengths. We found one subpopulation with intermediate waveforms that exhibits a sizable and significant correlation between target and movement tuning, suggesting that this subpopulation includes many cells that directly drive motor output.

Our modeling approach is predicated on the existence of a well defined, causal link between the preparatory activity in PMd/M1 and motor behavior, but it is agnostic to the particulars of this link. The true causal link may be the result of complex neural dynamics (Shenoy et al., 2013; Sussillo et al., 2015), the mechanics of the population decoding computation (Deneve et al., 1999; Jazayeri and Movshon, 2006; Hohll et al., 2013) or other gross network phenomena (Carandini and Heeger, 2011). We only assume that such intermediate dynamics can be well de-

scribed by a first-order perturbative approximation. Our model also focuses on the information about the target that flows through the recorded neural population; any independent influence of other neurons will simply act as a source of downstream noise. Nonetheless, the assumption of minimum-variance motor weightings allows us to predict properties of these measured neurons, despite being unable to record from all the relevant neurons in the circuit.

Several recent studies have drawn a distinction between “representational” and “dynamical” views of motor cortex (Churchland et al., 2010; Shenoy et al., 2013). We see two separate distinctions between these approaches: the distinction between causal and non-causal models of motor cortex and the distinction between static and dynamical models. Here, we have focused on the first of these, exploring the relationship between traditional, representational tuning-curve analyses and a simple causal model. We show that a static analysis of the mapping from delay-period activity to initial movement can help elucidate the role that the neurons play in transforming sensory input to motor output. A similar approach was used by Lillicrap and Scott (2013), who showed that a static approximation of their dynamic causal model of M1 movement generation could explain the anisotropic distribution of preferred directions observed in M1.

Comparison with other efforts to categorize neuronal function in PMd and M1

We have shown that computing tuning curves from center-out reaching data in the standard way (Georgopoulos et al., 1982) conflates target and movement tuning. Nonetheless, many groups have used center-out reaching combined with sophisticated experimental manipulations and analyses to distinguish neurons whose activity is more related to either task-relevant sensory cues or motor output. For example, sensory and motor activity have been dissociated by varying, independently from the spatial target cue, the presence of motor response (Wise et al., 1983) or the effector used (Cisek et al., 2003). Similar distinctions have been made based on the difference between movement and posture-related activity (Scott and Kalaska, 1995; Crammond and Kalaska, 1996). Also, many studies have used the time course of activity during the trial as a means of categorizing neural responses (Song and McPeck, 2010; Suminski et al., 2015). These previous studies focused on mean firing rates across trials, either in particular trial epochs or in peri-stimulus and peri-movement histograms. In contrast, we have focused on understanding how the trial-by-trial variability of neural activity relates to the trial-by-trial variability of behavior, which has allowed us to test causal models of how information flows through the PMd/M1 network (Fig. 6).

A key finding of our study is that neurons with intermediate waveform durations are most consistent, at a population level, with the predictions for output neurons. Although previous studies have identified cells with the shortest waveforms as putative inhibitory neurons and those with the longest waveforms as putative output cells (Mountcastle et al., 1969; Mitchell et al., 2007), we did not see corroborating evidence for this pattern in our data. Rather, these populations showed little or no correlations between the target and movement tuning, suggesting that they mostly consist of nonprojection neurons, or at least that they are heterogeneous. This is consistent with the results of Vigneswaran et al. (2011), who found primarily overlapping distributions of waveform lengths for identified pyramidal tract neurons and for nonidentified cells.

Two other groups have examined the relationship between the functional roles of individual neurons in the sensorimotor trans-

formation and their waveform shape. Song and McPeck (2010) took advantage of reaction time variability in reaching to categorize cells as being aligned to cue onset or to movement initiation. They found that cells with short waveforms ($<300 \mu\text{s}$) were primarily cue aligned (but see Kaufman et al., 2010), whereas longer waveform ($300\text{--}500 \mu\text{s}$) cells included a mix of cue- and movement-aligned cells. A very similar analysis was performed in macaque frontal eye fields by Cohen et al. (2009), who found a broad distribution of waveform lengths, similar to our own. They found that although putative output cells (those whose activity is saccade aligned) had a broad distribution of waveform lengths, the mode of that distribution was at intermediate lengths. Moreover, in remarkable agreement with our own results, they found a narrow bin of waveform lengths, centered at $370 \mu\text{s}$, which was composed almost entirely of putative output cells.

In addition to its inherent scientific interest, the electrophysiological identification of neurons by their functional/circuit roles may be of practical use. For example, in brain-machine interface applications, this might allow for the selection of a sub-population of cells that are more relevant neurons for decoding (Best et al., 2016). Furthermore, because output neurons may be more representative of what the animal “believes” to be the output of the circuit, use of these cells may allow for faster learning and easier brain-machine interface control.

References

- Best MD, Takahashi K, Suminski AJ, Ethier C, Miller LE, Hatsopoulos NG (2016) Comparing offline decoding performance in physiologically defined neuronal classes. *J Neural Eng* 13:026004. [CrossRef Medline](#)
- Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. *Nat Rev Neurosci* 13:51–62. [CrossRef Medline](#)
- Chaisanguanthum KS, Shen HH, Sabes PN (2014) Motor variability arises from a slow random walk in neural state. *J Neurosci* 34:12071–12080. [CrossRef](#)
- Churchland MM, Afshar A, Shenoy KV (2006) A central source of movement variability. *Neuron* 52:1085–1096. [CrossRef Medline](#)
- Churchland MM, Cunningham JP, Kaufman MT, Ryu SI, Shenoy KV (2010) Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron* 68:387–400. [CrossRef Medline](#)
- Cisek P, Crammond DJ, Kalaska JF (2003) Neural activity in primary motor and dorsal premotor cortex in reaching tasks with the contralateral versus ipsilateral arm. *J Neurophysiol* 89:922–942. [Medline](#)
- Cohen JY, Pouget P, Heitz RP, Woodman GF, Schall JD (2009) Biophysical support for functionally distinct cell types in the frontal eye field. *J Neurophysiol* 101:912–916. [Medline](#)
- Crammond DJ, Kalaska JF (1996) Differential relation of discharge in primary motor cortex and premotor cortex to movements versus actively maintained postures during a reaching task. *Exp Brain Res* 108:45–61. [Medline](#)
- Deneve S, Latham PE, Pouget A (1999) Reading population codes: a neural implementation of ideal observers. *Nat Neurosci* 2:740–745. [CrossRef Medline](#)
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433. [CrossRef Medline](#)
- Georgopoulos AP, Carpenter AF (2015) Coding of movements in the motor cortex. *Curr Opin Neurobiol* 33:34–39. [CrossRef Medline](#)
- Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J Neurosci* 2:1527–1537. [Medline](#)
- Georgopoulos AP, Kettner RE, Schwartz AB (1988) Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *J Neurosci* 8:2928–2937. [Medline](#)
- Gold C, Henze DA, Koch C, Buzsáki G (2006) On the origin of the extracellular action potential waveform: a modeling study. *J Neurophysiol* 95:3113–3128. [CrossRef Medline](#)
- Gordon J, Ghilardi MF, Cooper SE, Ghez C (1994) Accuracy of planar

- reaching movements. II. Systematic extent errors resulting from inertial anisotropy. *Exp Brain Res* 99:112–130. [CrossRef Medline](#)
- Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. *Nature* 394:780–784. [CrossRef Medline](#)
- Hohl SS, Chaisanguanthum KS, Lisberger SG (2013) Sensory population decoding for visually guided movements. *Neuron* 79:167–179. [CrossRef Medline](#)
- Huang X, Lisberger SG (2009) Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J Neurophysiol* 101:3012–3030. [CrossRef Medline](#)
- Huntley GW, Jones EG (1991) Relationship of intrinsic connections to forelimb movement representations in monkey motor cortex: a correlative anatomic and physiological study. *J Neurophysiol* 66:390–413. [Medline](#)
- Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–696. [CrossRef Medline](#)
- Jones EG, Wise SP (1977) Size, laminar and columnar distribution of efferent cells in the sensory-motor cortex of monkeys. *J Comp Neurol* 175:391–438. [CrossRef Medline](#)
- Kaufman MT, Churchland MM, Santhanam G, Yu BM, Afshar A, Ryu SJ, Shenoy KV (2010) Roles of monkey premotor neuron classes in movement preparation and execution. *J Neurophysiol* 104:799–810. [CrossRef Medline](#)
- Lillicrap TP, Scott SH (2013) Preference distributions of primary motor cortex neurons reflect control solutions optimized for limb biomechanics. *Neuron* 77:168–179. [CrossRef Medline](#)
- McGuire LM, Sabes PN (2011) Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. *J Neurosci* 31:6661–6673. [CrossRef](#)
- Mehring C, Rickert J, Vaadia E, Cardoso de Oliveira S, Aertsen A, Rotter S (2003) Inference of hand movements from local field potentials in monkey motor cortex. *Nat Neurosci* 6:1253–1254. [CrossRef Medline](#)
- Mehring C, Nawrot MP, de Oliveira SC, Vaadia E, Schulze-Bonhage A, Aertsen A, Ball T (2004) Comparing information about arm movement direction in single channels of local and epicortical field potentials from monkey and human motor cortex. *J Physiol (Paris)* 98:498–506. [CrossRef Medline](#)
- Merchant H, Naselaris T, Georgopoulos AP (2008) Dynamic sculpting of directional tuning in the primate motor cortex during three-dimensional reaching. *J Neurosci* 28:9164–9172. [CrossRef](#)
- Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55:131–141. [CrossRef Medline](#)
- Mountcastle VB, Talbot WH, Sakata H, Hyvärinen J (1969) Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys. Neuronal periodicity and frequency discrimination. *J Neurophysiol* 32:452–484. [Medline](#)
- Paninski L, Fellows MR, Hatsopoulos NG, Donoghue JP (2004) Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J Neurophysiol* 91:515–532. [Medline](#)
- Quian Quiroga R (2009) What is the real shape of extracellular spikes? *J Neurosci Methods* 177:194–198. [CrossRef](#)
- Sanger TD (1996) Probability density estimation for the interpretation of neural population codes. *J Neurophysiol* 76:2790–2793. [Medline](#)
- Schwartz AB (2004) Cortical neural prosthetics. *Annu Rev Neurosci* 27:487–507. [CrossRef Medline](#)
- Schwartz AB, Moran DW (2000) Arm trajectory and representation of movement processing in motor cortical activity. *Eur J Neurosci* 12:1851–1856. [CrossRef Medline](#)
- Schwartz AB, Kettner RE, Georgopoulos AP (1988) Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. Relations between single cell discharge and direction of movement. *J Neurosci* 8:2913–2927. [Medline](#)
- Scott SH, Kalaska JF (1995) Changes in motor cortex activity during reaching movements with similar hand paths but different arm postures. *J Neurophysiol* 73:2563–2567. [Medline](#)
- Shenoy KV, Sahani M, Churchland MM (2013) Cortical control of arm movements: a dynamical systems perspective. *Annu Rev Neurosci* 36:337–359. [CrossRef Medline](#)
- Shoham S, Fellows MR, Normann RA (2003) Robust, automatic spike sorting using mixtures of multivariate t-distributions. *J Neurosci Methods* 127:111–122. [CrossRef](#)
- Song JH, McPeck RM (2010) Roles of narrow- and broad-spiking dorsal premotor area neurons in reach target selection and movement production. *J Neurophysiol* 103:2124–2138. [CrossRef Medline](#)
- Suminski AJ, Mardoum P, Lillicrap TP, Hatsopoulos NG (2015) Temporal evolution of both premotor and motor cortical tuning properties reflect changes in limb biomechanics. *J Neurophysiol* 113:2812–2823. [CrossRef Medline](#)
- Sussillo D, Churchland MM, Kaufman MT, Shenoy KV (2015) A neural network that finds a naturalistic solution for the production of muscle activity. *Nat Neurosci* 18:1025–1033. [CrossRef Medline](#)
- Todorov E (2000) Direct cortical control of muscle activation in voluntary arm movements: a model. *Nat Neurosci* 3:391–398. [CrossRef Medline](#)
- Todorov E (2004) Optimality principles in sensorimotor control. *Nat Neurosci* 7:907–915. [CrossRef Medline](#)
- Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. *Nat Neurosci* 5:1226–1235. [CrossRef Medline](#)
- Tokuno H, Nambu A (2000) Organization of nonprimary motor cortical inputs on pyramidal and nonpyramidal tract neurons of primary motor cortex: an electrophysiological study in the macaque monkey. *Cereb Cortex* 10:58–68. [CrossRef](#)
- van Beers RJ (2009) Motor learning is optimally tuned to the properties of motor noise. *Neuron* 63:406–417. [CrossRef Medline](#)
- Verstynen T, Sabes PN (2011) How each movement changes the next: an experimental and theoretical study of fast adaptive priors in reaching. *J Neurosci* 31:10050–10059. [CrossRef](#)
- Vigneswaran G, Kraskov A, Lemon RN (2011) Large identified pyramidal cells in macaque motor and premotor cortex exhibit “thin spikes”: implications for cell type classification. *J Neurosci* 31:14235–14242. [CrossRef](#)
- Wise SP, Weinrich M, Mauritz KH (1983) Motor aspects of cue-related neuronal activity in premotor cortex of the rhesus monkey. *Brain Res* 260:301–305. [CrossRef Medline](#)
- Yokoi I, Komatsu H (2010) Putative pyramidal neurons and interneurons in the monkey parietal cortex make different contributions to the performance of a visual grouping task. *J Neurophysiol* 104:1603–1611. [CrossRef Medline](#)