Article

Neuron

Motor Cortex Embeds Muscle-like Commands in an Untangled Population Response

Highlights

- Motor cortex displays a signature of a smooth dynamical system: low tangling
- Low tangling explains the previously puzzling dominant signals in motor cortex
- Low tangling confers noise robustness and predicts population activity patterns
- Motor cortex embeds output commands in structure that reduces tangling

Authors

Abigail A. Russo, Sean R. Bittner, Sean M. Perkins, ..., Laurence F. Abbott, John P. Cunningham, Mark M. Churchland

Correspondence

mc3502@columbia.edu

In Brief

Using a novel extended movement task, Russo et al. show that neural activity in motor cortex is dominated by non-muscle-like signals. A computational approach reveals that these dominant features are expected and can be predicted given the constraint that neural activity produces muscle commands while obeying a smooth flow-field.

Russo et al., 2018, Neuron *97*, 1–14
February 21, 2018
https://doi.org/10.1016/j.neuron.2018.01.004

CellPress

Please cite this article in press as: Russo et al., Motor Cortex Embeds Muscle-like Commands in an Untangled Population Response, Neuron (2018), https://doi.org/10.1016/j.neuron.2018.01.004

Neuron
Article

Motor Cortex Embeds Muscle-like Commands in an Untangled Population Response

Abigail A. Russo,[1,2] Sean R. Bittner,[1,2] Sean M. Perkins,[2,3] Jeffrey S. Seely,[1,2] Brian M. London,[4] Antonio H. Lara,[1,2] Andrew Miri,[1,2,7] Najja J. Marshall,[1,2] Adam Kohn,[8] Thomas M. Jessell,[1,2,5,6,7] Laurence F. Abbott,[1,2,5,9,11] John P. Cunningham,[2,10,11,12] and Mark M. Churchland[1,2,5,10,13,*]
[1]Department of Neuroscience, Columbia University Medical Center, New York, NY 10032, USA
[2]Zuckerman Institute, Columbia University, New York, NY 10027, USA
[3]Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA
[4]SeatGeek, New York, NY 10003, USA
[5]Kavli Institute for Brain Science, Columbia University Medical Center, New York, NY 10032, USA
[6]Howard Hughes Medical Institute, Columbia University, New York, NY 10032, USA
[7]Departments of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA
[8]Department of Ophthalmology and Visual Sciences, Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY 10461, USA
[9]Department of Physiology and Cellular Biophysics, Columbia University Medical Center, New York, NY 10032, USA
[10]Grossman Center for the Statistics of Mind, Columbia University, New York, NY 10027, USA
[11]Center for Theoretical Neuroscience, Columbia University Medical Center, New York, NY 10032, USA
[12]Department of Statistics, Columbia University, New York, NY 10027, USA
[13]Lead Contact
*Correspondence: mc3502@columbia.edu
https://doi.org/10.1016/j.neuron.2018.01.004

SUMMARY

Primate motor cortex projects to spinal interneurons and motoneurons, suggesting that motor cortex activity may be dominated by muscle-like commands. Observations during reaching lend support to this view, but evidence remains ambiguous and much debated. To provide a different perspective, we employed a novel behavioral paradigm that facilitates comparison between time-evolving neural and muscle activity. We found that single motor cortex neurons displayed many muscle-like properties, but the structure of population activity was not muscle-like. Unlike muscle activity, neural activity was structured to avoid "tangling": moments where similar activity patterns led to dissimilar future patterns. Avoidance of tangling was present across tasks and species. Network models revealed a potential reason for this consistent feature: low tangling confers noise robustness. Finally, we were able to predict motor cortex activity from muscle activity by leveraging the hypothesis that muscle-like commands are embedded in additional structure that yields low tangling.

INTRODUCTION

For 50 years, a central question in motor physiology has been whether motor cortex activity resembles muscle activity, and, if not, why not (Evarts, 1968)? Primate motor cortex is as close as one synapse to the motoneurons (Rathelot and Strick, 2009), and single action potentials in corticospinal neurons can measurably impact muscle activity (Cheney and Fetz, 1980; Schieber and Rivlis, 2007), suggesting that motor cortex may encode muscle-like commands (Ajemian et al., 2008; Herter et al., 2009; Morrow et al., 2009; Sergio et al., 2005; Todorov, 2000). Yet motor cortical responses often differ from patterns of muscle force, motivating the hypothesis that motor cortex might primarily encode movement velocity or direction (Georgopoulos et al., 1986; Moran and Schwartz, 1999b; Schwartz, 1994, 2007). Alternatively, it has been proposed that non-muscle-like response features reflect network or feedback dynamics (Churchland and Cunningham, 2014; Churchland et al., 2012; Kaufman et al., 2016; Lillicrap and Scott, 2013; Maier et al., 2005; Michaels et al., 2016; Rokni and Sompolinsky, 2012; Seely et al., 2016; Shenoy et al., 2013; Sussillo et al., 2015). Many studies, largely focused on reaching, have produced little consensus (Aflalo and Graziano, 2007; Fetz, 1992; Georgopoulos et al., 2007; Moran and Schwartz, 2000; Mussa-Ivaldi, 1988; Reimer and Hatsopoulos, 2009; Scott, 2008).

The ubiquity of reaching tasks has naturally promoted analysis of directional tuning (e.g., Ajemian et al., 2008; Georgopoulos et al., 1982; Kakei et al., 1999; Lillicrap and Scott, 2013; Scott and Kalaska, 1997), the interpretation of which remains debated (Georgopoulos et al., 2007; Moran and Schwartz, 2000; Mussa-Ivaldi, 1988; Sanger, 1994). More generally, reaching tasks tend to inspire hypotheses where neurons encode parameters relevant to reaching (Burnod et al., 1992; Georgopoulos et al., 1982, 1986; Moran and Schwartz, 1999b) or reflect reach-appropriate dynamics (Churchland and Cunningham, 2014; Churchland et al., 2012). A few studies (Hatsopoulos et al., 2007; Moran and Schwartz, 1999a; Schwartz et al., 2004) examined primate motor cortex during extended drawing or tracing movements but also focused largely on directional properties (although see Fitzsimmons et al., 2009; Foster et al., 2014). Given that the

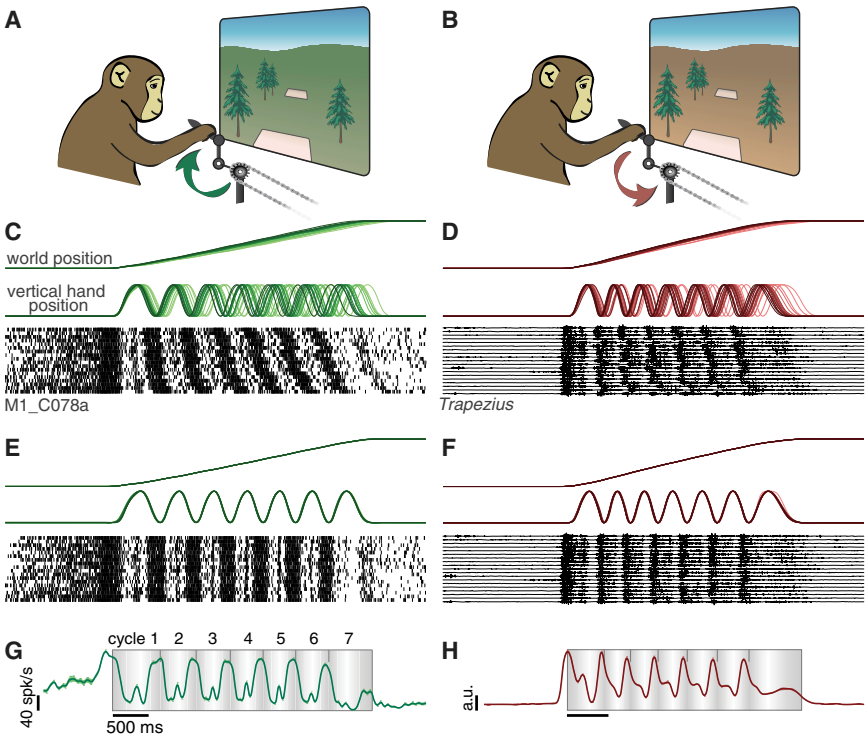

Figure 1. Behavioral and Physiological Responses during Cycling

(A) Schematic of the task during forward cycling. A green landscape indicated that virtual progress required cycling forward.

(B) An orange landscape indicated that progress required cycling backward.

(C) Behavioral data and spikes from one neuron during an example session. Data are for a single condition: forward/seven-cycle/bottom-start (monkey C). Trials are aligned to movement onset and ordered from fastest to slowest.

(D) Behavioral data and raw trapezius EMG for one condition: backward/seven-cycle/bottom-start (monkey D).

(E) Data from (C) after temporal scaling to align trials.

(F) Data from (D) after temporal scaling.

(G) Trial-averaged and filtered neural activity for the example neuron in (C) and (E). Envelopes show the standard error of the mean (SEM), which was often within the trace width. Shading tracks vertical hand position: lightest at top and darkest at bottom. Small tick marks indicate each cycle's completion.

(H) Rectified, filtered, and trial-averaged EMG for the example in (D) and (F).

defining feature of movement is change with time, progress may benefit from more detailed comparisons of time-evolving patterns of neural and muscle activity. To afford such comparisons, an ideal task would achieve the traditional goal of dissociating kinematics from muscle activity (Kakei et al., 1999; Scott and Kalaska, 1997), but in the temporal rather than spatial domain. This has been achieved during reaches (Churchland and Shenoy, 2007; Sergio et al., 2005), but more extended movements could improve the power of such comparisons.

Unlike in sensory systems where responses strongly reflect incoming stimuli, time-evolving responses in the motor system may reflect computations performed by internal and feedback dynamics. A growing body of work seeks to understand neural responses in terms of signals that a recurrent or feedback-driven neural network would need to perform the relevant task (Hennequin et al., 2014; Li et al., 2016; Lillicrap and Scott, 2013; Mante et al., 2013; Michaels et al., 2016; Sussillo and Barak, 2013). Although multiple network solutions are typically possible, broad principles can still apply. For example, the simple constraint of a smooth dynamical flow-field explains aspects of neural dynamics during reaching (Sussillo et al., 2015).

Here, we leverage a "cycling" task that evoked extended movements with simple kinematics driven by temporally complex patterns of muscle activity. We found that single neurons and muscles shared many temporal response properties. Yet the neural population as a whole was dominated by signals that were not muscle-like and were not explained by velocity/direction coding. Seeking an alternative explanation, we focused on a basic principle of recurrent and feedback-driven networks: the present network state strongly influences the future state. Thus, two similar patterns of activity, observed at different moments, should not lead to highly dissimilar patterns in the near future. We refer to violations of this principle as "trajectory tangling." Moments of high tangling imply either a potential instability in network dynamics or a moment when the system must rely on external commands.

Tangling was often high for muscle population trajectories. This was expected: muscles reflect descending commands and need not avoid tangling. In contrast, tangling was very low for motor cortex population trajectories. This was found not only during cycling, but also during a reaching task, and in rodent during reach-to-grasp and locomotion. However, low tangling was anatomically specific and was not observed for primary visual or somatosensory cortex. We found that the dominant signals in motor cortex were those that naturally reduced tangling. Using an optimization approach, we could quantitatively predict the neural population response based on only two principles: the need to encode muscle-like commands and the need to ensure low tangling. Network simulations confirm that low trajectory tangling is computationally beneficial. Networks with lower tangling are more noise robust. In summary, our data reveal a potentially general property of motor cortex: muscle-like signals are present but are relatively modest "ripples" riding on top of larger signals that confer minimal tangling. Thus, the dominant signals in motor cortex may serve not a representational function—encoding specific variables—but rather a computational function: ensuring that outgoing commands can be generated reliably.

RESULTS

Task and Behavior

We trained two rhesus macaque monkeys to grasp a hand-pedal and cycle for juice reward. Cycling produced movement

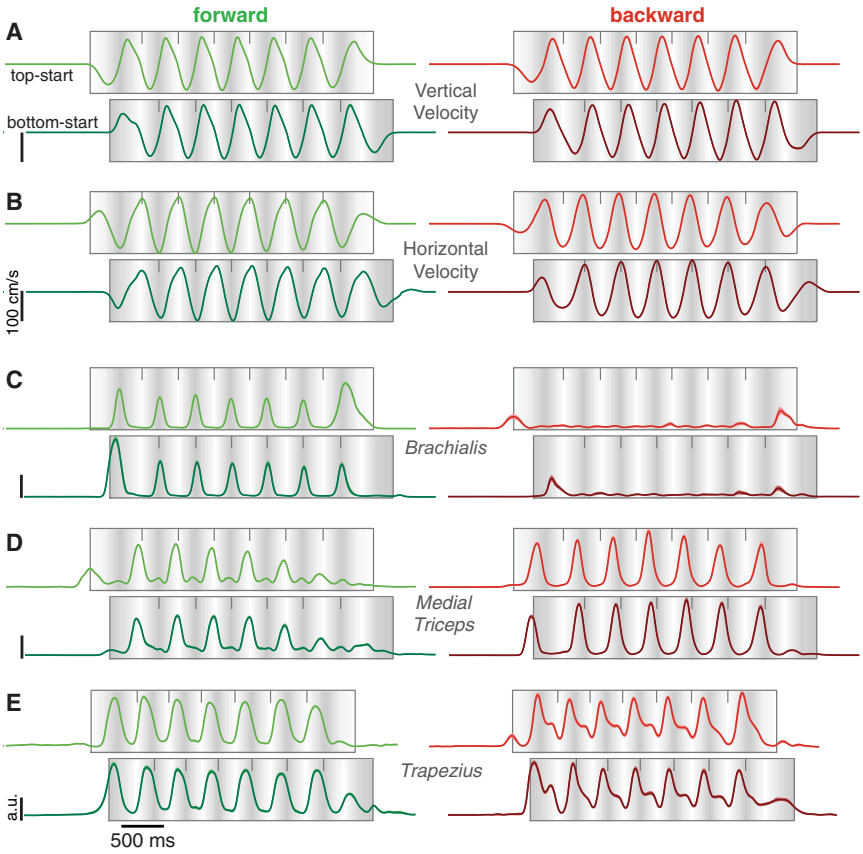

Figure 2. Kinematics and Muscle Activity

(A) Vertical hand velocity, averaged across trials from a typical session (monkey C). Format as in Figure 1G. Left and right columns show data for forward and backward seven-cycle movements. Data for top- and bottom-start movements are shifted to align hand position (light shading indicates cycle apex).

(B) Corresponding horizontal hand velocity traces.

(C) Brachialis EMG (monkey C). Envelopes show SEM.

(D) Medial triceps EMG (monkey C).

(E) Trapezius EMG (monkey D).

through a virtual landscape. Landscape color indicated whether forward virtual motion required "forward" cycling (Figure 1A) or "backward" cycling (Figure 1B). During each trial, the monkey progressed from one stationary target to another. Target acquisition required a stationary pedal with the target "under" the first person perspective (Figures 1A and 1B). The first target was acquired with a pedal orientation either straight up ("top-start") or straight down ("bottom-start"). Inter-target distance determined the required number of revolutions: 0.5, 1, 2, 4, or 7 cycles. Monkeys performed all combinations of two cycling directions, two starting orientations, and five distances. Cycling required overcoming simulated inertia and viscosity while countering the weight of an arm extended in front of the body. These requirements differ from those during locomotion and had to be learned.

Behavior was highly stereotyped; note similarity of virtual-world-position traces across trials in Figures 1C and 1D. Nevertheless, small trial-to-trial variations in cycling speed caused accumulating misalignment of kinematics with time. We thus temporally scaled trials so that virtual-world-position traces were closely matched. Doing so revealed considerable temporal structure in neural and electromyographic (EMG) responses (Figures 1E and 1F). To summarize such structure, we computed average firing rate (Figure 1G) or muscle activation (Figure 1H) across trials. We used a narrow filter (25-ms Gaussian kernel) relative to the timescale of behavior (~500-ms cycling period) to preserve fine temporal features. We similarly computed trial-averaged responses for key kinematic parameters such as hand velocity. Consistent with the circular motion, vertical and horizontal hand velocity exhibited approximately sinusoidal profiles (Figures 2A and 2B) that repeated across middle cycles and were slightly slower during initial/terminal cycles as angular velocity ramped up and down. Top- and bottom-start movements differed in phase but were otherwise similar during middle cycles.

Intramuscular EMG recordings (35 and 29 sites in monkey D and C) concentrated on muscles that moved the shoulder and elbow and to a lesser degree the wrist (which had limited mobility given the pedal design). Muscle activity (Figures 2C–2E) generally followed intuitions from biomechanics. For example, the triceps muscle extends the elbow, moving the hand away from the body. Accordingly, triceps activity (Figure 2D) peaked near each cycle's apex (white shading) when cycling forward and near its bottom (dark shading) when cycling backward. Some muscle responses were roughly sinusoidal and resembled kinematics, yet deviations from sinusoidal were common (e.g., Figure 2E).

Single-Neuron Responses

Well-isolated single neurons (103 and 109, monkeys D and C) were sequentially recorded from motor cortex, including sulcal and surface primary motor cortex and the immediately adjacent aspect of dorsal premotor cortex (potential differences within this population are explored later). Recordings were localized to the region where microstimulation activated muscles from which we recorded. Cycling evoked strong responses; nearly all neurons that could be isolated were task modulated. Peak firing rates ranged from 16 to 184 spikes/s (monkey D, mean: 69 spikes/s) and 16 to 185 spikes/s (monkey C, mean: 76 spikes/s). Neurons displayed a variety of intricate response patterns (Figure 3). These patterns were statistically reliable: SEMs were small and the same pattern could be seen repeatedly across middle cycles for both top- and bottom-start conditions.

Inspection revealed three features shared between muscles and neurons. First, responses often deviated from the sinusoidal profile of kinematics (e.g., Figure 2E, backward; Figure 3A, forward). Second, responses during initial/terminal cycles often

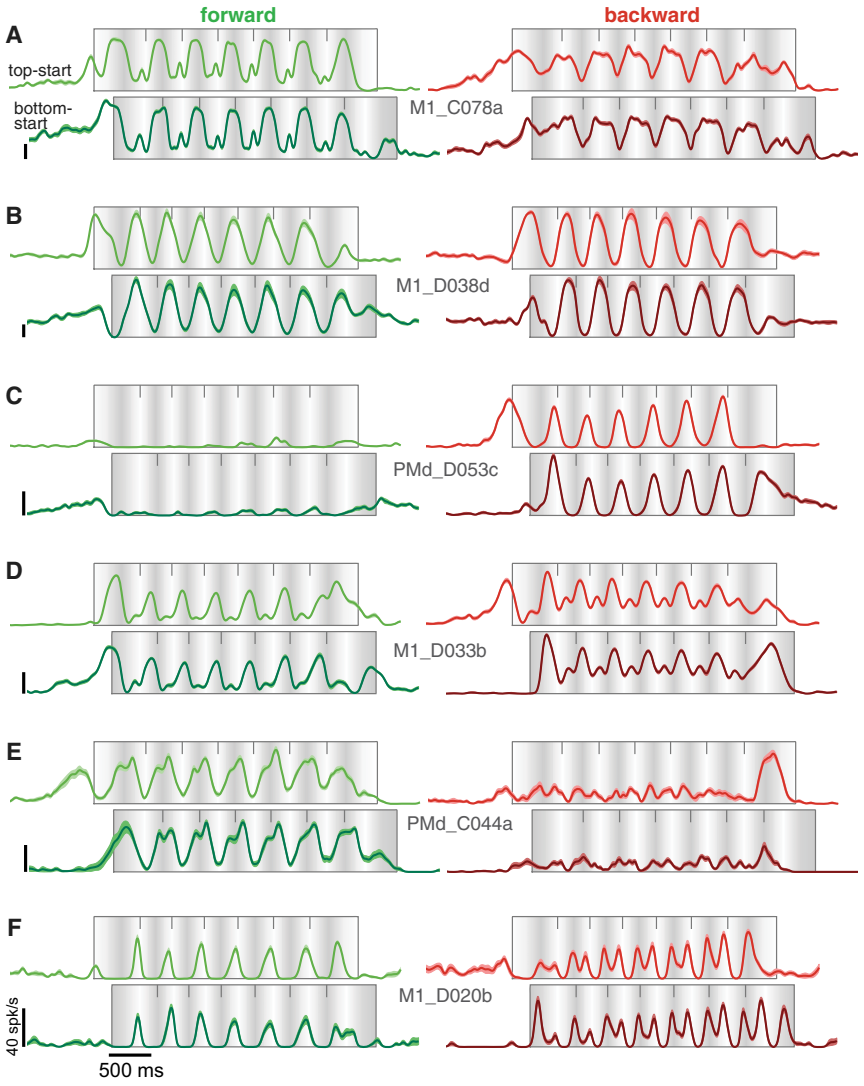

Figure 3. Responses of Six Example Motor Cortex Neurons

Format as for Figure 2.

(A–F) Average firing rate was computed across a median of 15 trials/condition per neuron. Neuron names indicate primary motor cortex (M1) versus dorsal premotor cortex (PMd) and monkey (D versus C). Calibrations are 40 spikes/s.

displayed differences in amplitude or temporal profile compared to middle cycles (e.g., Figure 2D, forward; Figure 3D, forward; Figure 3E, backward). This effect presumably relates to the unique force patterns required to start and stop. Third, responses could differ between forward and backward cycling in both amplitude (e.g., Figures 2C and 3C) and structure (e.g., Figures 2E, 3A, and 3F).

Consistent with these shared features, muscle responses could be successfully decoded from the neural population using a linear model (leave-one-out-cross-validated R^2 = 0.80 and 0.78) consistent with prior studies (Griffin et al., 2008; Morrow et al., 2009; Schieber and Rivlis, 2007). This is potentially impressive, given that a linear model is almost certainly too simplistic. This finding might suggest that motor cortex activity primarily reflects muscle-like commands. However, decoding neural activity from muscle activity was less successful (leave-one-out-cross-validated R^2 = 0.54 and 0.50). This discrepancy in fit quality was not simply due to neural recordings having higher sampling error than muscle recordings. The same discrepancy was observed when neural responses were de-noised using dimensionality reduction techniques (STAR Methods). Thus, while muscle-like signals can be found in the neural data, there exist additional, non-muscle-like neural response patterns.

Non-muscle-like Signals Dominate the Neural Population Response

To characterize population responses, we applied principal component analysis (PCA), a standard unsupervised algorithm that identifies the dominant signals in multi-dimensional data (Figure 4). Each such signal is a weighted combination of individual-neuron responses, with those weights (the PCs) optimized such that a small number of signals faithfully summarizes the full population response. We first examine the signals captured by the top two PCs. Plotting these signals versus one another yields a state-space trajectory (Figure 4C). Each point on the trajectory (e.g., the orange dot in Figure 4C) corresponds to the neural state at one moment (dashed line in Figures 4A and 4B). A two-dimensional trajectory provides only a partial summary of the neural state, but the resulting visualization can still be informative and inspire hypotheses.

Neural trajectories for monkey D are shown during both forward and backward cycling (Figure 4E, top and bottom subpanels). Top-start and bottom-start trajectories are superimposed. For monkey C, trajectories during forward and backward cycling are also superimposed (Figure 4H). For illustrative purposes, data are shown only for seven-cycle conditions (as in Figures 1, 2, and 3). Middle cycles (3–5) are highlighted in color. Neural trajectories followed repeating orbits throughout the middle cycles. Rotating orbits are expected during cycling, in contrast to reaching (Churchland et al., 2012), and simply reflect what can be observed in single neurons: middle-cycle responses tend to repeat. Muscle trajectories also followed repeating orbits (Figures 4D and 4G). Despite this basic similarity, neural and muscle trajectories behaved differently. Muscle trajectories counter-rotated: they orbited in opposing directions for forward and backward cycling. Counter-rotation is expected given the reversal of required force patterns. For example, forward cycling

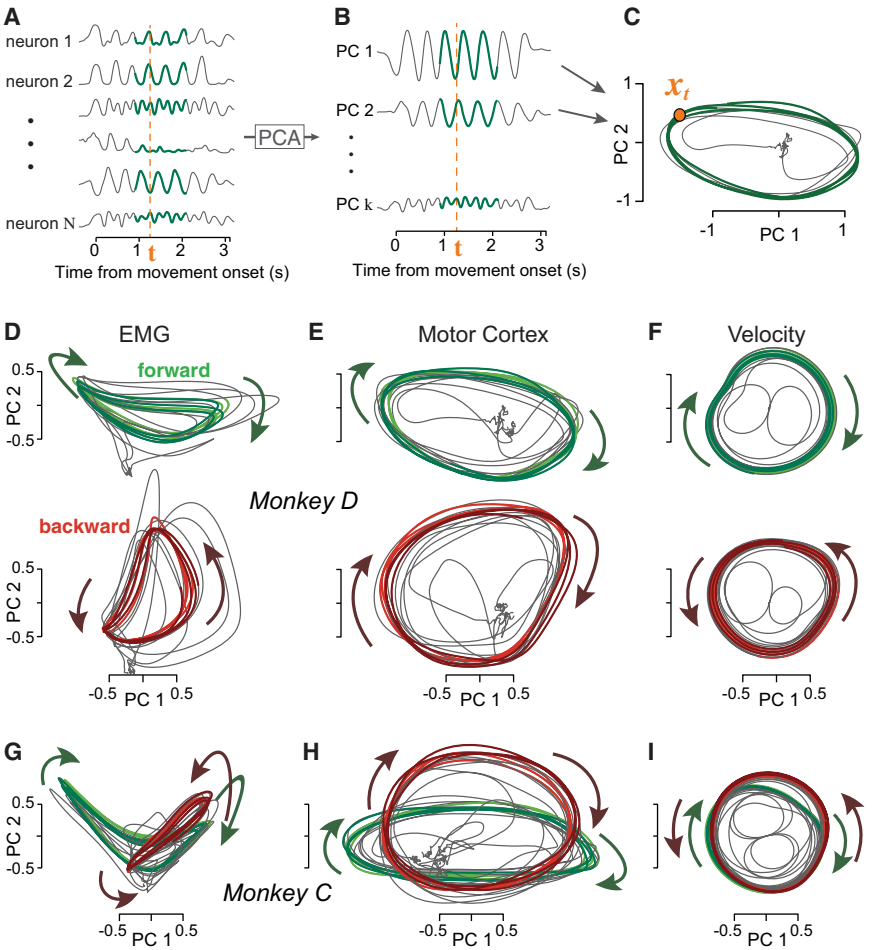

Figure 4. Visualization of Population Structure via PCA

(A) PCA operates on a population of responses (6 of 103 neurons are shown). Green traces highlight the middle cycles used to find the PCs for this visualization (subsequent analyses consider all times). PCs were computed based on cycling in both directions and both starting positions. Data are plotted only for the forward, bottom-start condition.

(B) Projections onto the PCs. The neural state at a given time (orange line) can be summarized by the values of the projections at that time.

(C) Corresponding neural trajectory. The projection onto the second PC is plotted against that onto the first (~35% of variance is captured in these dimensions). Orange dot shows the neural state at the same time as in (A) and (B).

(D) Muscle trajectories captured by projecting the muscle population response onto its first two PCs (monkey D). Trajectories are shown for forward and backward cycling, using the same PCs. Trajectories for top- and bottom-start conditions (lighter and darker colored traces, respectively) are overlaid.

(E) Corresponding neural trajectories.

(F) Corresponding hand-velocity trajectories, produced by applying PCA to horizontal and vertical velocity across multiple sessions. This is similar (but for a change of axes) to plotting average vertical versus horizontal velocity.

(G–I) PCA-based muscle, neural, and velocity trajectories for monkey C. Same format as (D)–(F), but trajectories for forward and backward cycling are overlaid.

requires lifting before pushing and backward cycling requires pushing before lifting. In contrast, neural trajectories co-rotated: they orbited in the same direction for forward and backward cycling. Furthermore, muscle trajectories tended to depart from circular: the orbit often possessed a kidney- or saddle-like shape. In contrast, neural trajectories were more circular or elliptical. Thus, the dominant signals in the neural population differ from those in the muscle population.

Potential Explanations and Caveats

A potential explanation for non-muscle-like patterns in motor cortex is that they encode directional signals such as hand velocity (e.g., Moran and Schwartz, 1999b). This explanation initially seems appealing given the present data. For example, the neural trajectory during backward cycling for monkey D (Figure 4E, bottom) visually resembles the corresponding velocity trajectory (Figure 4F, bottom). However, velocity trajectories necessarily counter-rotate between forward and backward cycling (the same would be true of hand direction or position). The dominant signals in the neural data do just the opposite. Combined with the fact that single-neuron response profiles typically do not resemble hand velocity or position traces, it seems unlikely that a simple representation of kinematic parameters can explain the dominant neural signals.

An alternative explanation is that the dominant neural signals may constitute descending commands to the muscles, yet may look non-muscle-like because they will be heavily modified by spinal circuitry. Cortical commands are likely integrated/low-pass filtered by the spinal cord (Shalit et al., 2012) and may encode muscle synergies rather than individual-muscle activations (Hart and Giszter, 2010). However, any commands related to force are almost certain to reverse between forward and backward cycling due to the reversal of required force patterns. Thus, the dominant signals in the neural data are not readily explained in terms of either muscle-command encoding or kinematic encoding. Of course, this does not rule out the possibility that muscle-like commands (or kinematic commands) are encoded in dimensions beyond the top two PCs. Indeed, we will suggest below that muscle-like commands likely are encoded. Yet, one is tempted to question the assumption that the dominant signals encode commands of any sort. Might there exist an alternative explanation?

Smooth Dynamics Predict Low Trajectory Tangling

Recent physiological and theoretical investigations suggest that the neural state in motor cortex obeys smooth dynamics (Churchland et al., 2012; Hall et al., 2014; Michaels et al., 2016; Seely et al., 2016; Sussillo et al., 2015). Smooth dynamics imply that neural trajectories should not be tangled: similar neural

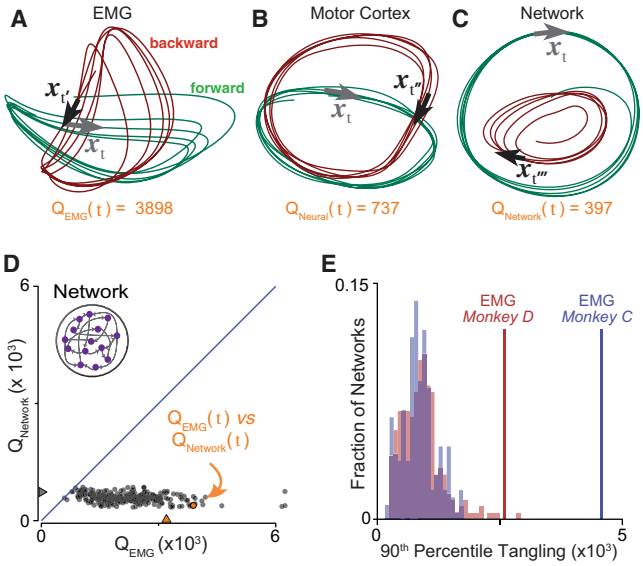

Figure 5. Illustration of the Trajectory Tangling Metric

(A) Muscle trajectories during the middle five cycles (of seven) for forward and backward cycling (bottom-start). Arrows illustrate a pair of states and their derivative (the trajectory direction). Time t resulted in a large $Q_{EMG}(t)$. Time t' is the "associated time" that resulted in that tangling value—i.e., that maximizes $\|\dot{\boldsymbol{x}}_t - \dot{\boldsymbol{x}}_{t'}\|^2 / \|\boldsymbol{x}_t - \boldsymbol{x}_{t'}\|^2 + \varepsilon$. In this example, t and t' occur during different conditions (forward versus backward). Tangling was computed in eight dimensions.

(B) Corresponding neural trajectories. Time t is the same as in A, and time t'' is the associated time that resulted in $Q_{Neural}(t)$.

(C) Corresponding trajectories from an artificial recurrent network, trained to produce the middle-cycle activity of all muscles.

(D) Scatterplot of network- versus muscle-trajectory tangling. One point per time/condition.

(E) Summary of tangling across 463 networks, each trained to produce the pattern of muscle activity from monkey D (red) or C (blue). For each network, we computed the 90th percentile tangling value across times/conditions. This distribution (across networks) can be compared to 90th percentile tangling for the empirical muscle populations (vertical lines).

states, either during different movements or at different times for the same movement, should not be associated with different derivatives. We quantified trajectory tangling using

$$Q(t) = \max_{t'} \frac{\|\dot{\boldsymbol{x}}_t - \dot{\boldsymbol{x}}_{t'}\|^2}{\|\boldsymbol{x}_t - \boldsymbol{x}_{t'}\|^2 + \varepsilon}, \qquad \text{(Equation 1)}$$

where $\boldsymbol{x}_t$ is the neural state at time t (i.e., a vector containing the neural responses at that time), $\dot{\boldsymbol{x}}_t$ is the temporal derivative of the neural state, $\|\cdot\|$ is the Euclidean norm, and ε is a small constant that prevents division by zero (STAR Methods). $Q(t)$ becomes high if there exists a state at a different time, t', that is similar but associated with a dissimilar derivative. We take the maximum to ask whether the state at time t ever becomes tangled with any other state. This maximum is taken with t' indexing across time during all conditions. $Q(t)$ can be analogously assessed for muscle trajectories.

We chose tangling as a straightforward measure of whether a given trajectory could have been produced by a smooth dynamical flow-field. Given limits on how non-smooth dynamics can be, moments of very high tangling are incompatible with a fixed flow-field. Furthermore, even moderately high tangling implies potential instabilities in the underlying flow-field (STAR Methods; Figure S1). High tangling thus implies either that the system must rely on external commands rather than internal dynamics, or that the system is flirting with instability. Although other metrics are possible, tangling has the practical benefit that it can be computed directly from the empirical trajectories without needing to know (or fit) a flow-field.

For the reasons above, a network that relies heavily on intrinsic dynamics should avoid tangling. In contrast, when population activity primarily reflects external commands (as for the muscles or a population of sensory neurons) high tangling is both benign and potentially necessary. For example, co-contraction of the biceps and triceps at one moment might need to be quickly followed by biceps activation and triceps relaxation. At a later moment or during a different movement, co-contraction might instead need to be followed by biceps relaxation and triceps activation. This would constitute an instance of tangling because the same state (co-contraction) is followed by different subsequent states. Do such moments of high tangling indeed occur for the muscles? If so, are they mirrored or avoided in the neural responses?

The state at a given time is a location on a state-space trajectory. The derivative is the direction in which the trajectory is headed. Two states are thus tangled if they are nearby but associated with different trajectory directions. For visualization, we consider a subset of the data: the middle five cycles of seven-cycle movements projected onto two dimensions (Figures 5A and 5B). Of course, two-dimensional projections only partially reflect the true population state; activity spans multiple dimensions. As a practical choice, we computed tangling in eight dimensions (results were robust with respect to this choice—see below). Muscle trajectories (Figure 5A) show three features suggestive of high tangling. First, muscle trajectories counter-rotate when cycling forward versus backward, yielding opposing derivatives for similar states. Second, muscle trajectories often cross themselves at right angles, resulting in similar states with very different derivatives. Third, non-circular trajectories sometimes yield nearby muscle states that move in different directions. These features indeed produced occasional moments of high tangling. For example, the gray arrow shows the muscle state and its derivative at a chosen time t. At time t', there exists another state at a similar location in state space but with a very different derivative (black arrow).

Neural trajectories (Figure 5B) appear potentially less tangled. Co-rotation prevents trajectories from continuously opposing one another between forward and backward cycling. Even within a condition, trajectories are closer to circular with fewer sharp bends. There are moments when trajectories cross in these two dimensions, but this did not result in high tangling because trajectories were separated in other dimensions. Notably, at moments when muscle trajectories became highly tangled, neural trajectories did not. For example, the muscle state at time t was strongly tangled, while the neural state at that same time was much less tangled.

Before comparing tangling across all times/conditions, we wished to confirm that the tangling metric behaves as intended when the ground truth is known. We examined trajectories

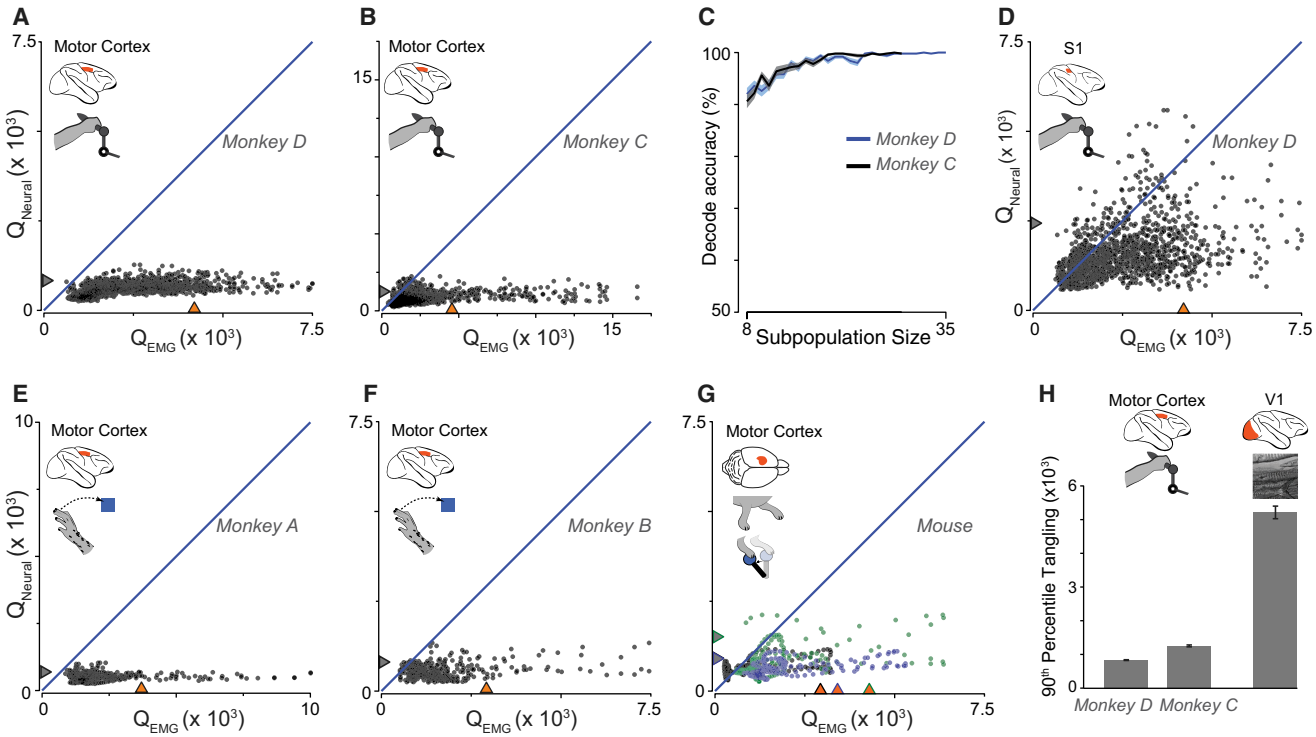

Figure 6. Trajectory Tangling for Multiple Datasets

(A) Motor cortex versus muscle trajectory tangling (monkey D). Points are shown for all times during movement for all twenty conditions. Blue line indicates unity slope. Gray/orange triangles denote 90th percentile tangling.

(B) Same for monkey C.

(C) Neural versus muscle populations are distinguishable based on tangling. For a given subpopulation size, we drew that many neurons and muscles and computed tangling. 500 such draws were made per subpopulation size. The vertical axis gives the percentage of instances where the neural subpopulation was correctly identified based on lower tangling. Envelopes show SEMs based on binomial statistics.

(D) S1 versus muscle trajectory tangling.

(E) Motor cortex versus muscle trajectory tangling during reaching (monkey A).

(F) Same but for monkey B.

(G) Motor cortex versus muscle trajectory tangling in three mice (black, blue, and green symbols). (Illustration by E. Daubert).

(H) Comparison of motor cortex and V1 trajectory tangling. Because V1 data have no corresponding muscle activity, tangling is quantified by 90th percentile values. Motor cortex data are from the cycling task as in (A) and (B). SEMs were computed via bootstrap: the distribution of tangling values was resampled 200 times, producing a sampling distribution of 90th percentile tangling values.

See also Figures S2 and S3.

from a simulated recurrent neural network trained to produce muscle activity for the subset of data plotted in Figure 5A. The network output closely resembled those muscle signals, yet the dominant signals internal to the network did not (cf. Figure 5C with Figure 5A). To compare tangling, we plotted $Q_{Network}(t)$ versus $Q_{EMG}(t)$ for every time during both simulated conditions (Figure 5D). Network-trajectory tangling was consistently lower than muscle-trajectory tangling, despite producing muscle trajectories as an output. We repeated this analysis for multiple simulated networks, using different weight initializations and meta-parameters. The degree of network-trajectory tangling was variable (distributions in Figure 5E) but was nearly always lower than muscle-trajectory tangling.

Neural- versus Muscle-Trajectory Tangling

For motor cortex, we compared Q_{Neural} and Q_{EMG} for all times across all twenty conditions. At least four results are possible. First, if motor cortex activity is a straightforward code for muscle activity, Q_{Neural} and Q_{EMG} should have a linear relationship with a slope near unity. Second, if motor cortex reflects unknown variables, and/or if tangling captures nothing fundamental, Q_{Neural} and Q_{EMG} may show no clear relationship. Third, if neural activity is more complex, intricate, or "noisier" than muscle activity, Q_{Neural} could tend to be greater than Q_{EMG}. Finally, Q_{Neural} could be systematically reduced relative to Q_{EMG}, as was the case for the simulated networks.

The data obeyed the final prediction (Figures 6A and 6B). The neural state was less tangled than the corresponding muscle state in 99.9% and 96.6% of cases (monkey D and C). The rare exceptions occurred when tangling was low for both. Strikingly, muscle-trajectory tangling could be quite high with no accompanying increase in neural-trajectory tangling. Statistically, distributions of Q_{Neural} and Q_{EMG} were indeed different (paired t test, $p < 10^{-10}$ for each monkey). The difference in tangling

was robust to analysis choices: it did not depend on the use of PCA versus "raw" data (Figure S2), on the number of PCs analyzed (Figure S3), on whether we matched dimensionality or variance explained (Figure S3), or on the relative number of neurons versus muscles (Figure S3). The large difference between Q_{Neural} and Q_{EMG} is striking given that visual inspection does not readily reveal whether individual recordings are neural or muscular (cf. Figure 3 with Figure 2). Yet the tangling metric readily distinguished between even small populations of neurons versus muscles (Figure 6C).

Tangling across Tasks, Species, and Areas

Is low neural- versus muscle-trajectory tangling specific to cycling or a more general property of motor cortex? We leveraged recently collected data (Elsayed et al., 2016) from two monkeys performing a center-out reach task. Again, Q_{Neural} was greatly reduced relative to Q_{EMG} (Figures 6E and 6F). We also compared Q_{Neural} and Q_{EMG} in mice during an experiment with two behaviors: reaching to pull a joystick and walking on a treadmill (Miri et al., 2017). We observed a slightly weaker yet similar effect (Figure 6G) to that seen in primates. Thus, low trajectory tangling in motor cortex appears to be a general property.

We also examined responses in the proprioceptive region (area 3a) of primary somatosensory cortex (S1) during cycling. This region is immediately adjacent to motor cortex, and individual-neuron responses (Figure S4) are surprisingly similar to those in motor cortex. Yet tangling was not as consistently low in S1 (Figure 6D) as it was in motor cortex (Figure 6A, same task and monkey). At moments where the muscle state became highly tangled, the S1 state often also became quite tangled. All three tangling distributions were significantly different: $p < 10^{-10}$ comparing muscle and S1 populations; $p < 10^{-10}$ comparing S1 and motor cortex populations (paired t test).

We also considered a primary visual cortex (V1) population responding to natural-scene movies. V1 trajectories were much more tangled than motor cortex trajectories (Figure 6H; $p < 10^{-10}$ and $p < 10^{-10}$, two-sample t test comparing V1 with motor cortex for monkey D and C). Across datasets (motor cortex, muscles, S1, V1), there was no clear relationship between dimensionality and tangling (Figure S5). Instead, tangling was highest for those populations (muscles and sensory areas) where driving inputs are expected to have the largest impact. This agrees with the fact that driving inputs, unless they can be predicted from outgoing commands, can readily cause the same state to be followed by different future states. For example, no constraint prevents image A from being followed by image B on one occasion, and by image C on another occasion.

Noise-Robust Networks Display Low Tangling

For a recurrent or feedback-driven network, it is intuitive that high tangling must be avoided. If the flow-field has some degree of smoothness, nearby states cannot be associated with very different derivatives. Thus, moments of high tangling cannot be produced without relying on disambiguating external inputs. Yet motor cortex trajectories avoided even moderate tangling. This is not strictly necessary even in the idealized case of a fully autonomous dynamical system. For example, some recurrent networks did show moderate tangling (right tail of the distribution in Figure 5E) yet still functioned. Might the very low empirical tangling confer some computational advantage? Formal considerations suggest so: even moderate tangling implies potential dynamical instabilities (STAR Methods).

To explore potential advantages of low tangling, we considered neural networks trained to generate a simple idealized output: cos t for one muscle and sin $2t$ for a second muscle (Figure 7A, top). The output trajectory was thus a figure eight (left subpanel). It is not possible for a network's internal trajectory to follow a pure figure eight; the center-most state is very highly tangled. Tangling can be reduced by employing a third dimension such that the trajectory is $[\cos t;\ \sin 2t;\ \beta \sin t]$. Even a modest value of β reduces tangling enough (middle subpanel) that the trajectory can be produced. As the network follows that three-dimensional trajectory, the figure-eight pattern can still be "read out" via projection, with the third dimension falling in the null space of the readout (Druckmann and Chklovskii, 2012; Kaufman et al., 2014). Are further decreases in tangling (right subpanel) advantageous? We examined noise tolerance across networks with internal trajectories $[\cos t;\ \sin 2t;\ \beta \sin t]$ and different values of β. This necessitated the unusual step of training networks not only to produce a desired output, but to follow a specified internal trajectory (STAR Methods).

Networks with high trajectory tangling failed to produce the figure-eight output in the presence of even small amounts of noise (Figure 7B). Networks with low trajectory tangling were much more noise robust. We performed a similar analysis with trajectories that encoded the empirical muscle trajectories, but with varying degrees of tangling (found using the optimization approach in the next section). Again, low tangling provided noise robustness (Figure S6). This was true both for networks that generated a single internal trajectory, and networks that generated different forward and backward trajectories based on inputs. Intuitively, when tangling is low, noise is less likely to perturb the network onto a nearby but inappropriate part of the trajectory. More formally, low tangling aids local stability (Figure S1; STAR Methods).

While the example in Figures 7A and 7B is simplified, it illustrates a feature that may help interpret the empirical neural trajectories. Setting $\beta = 1$ yields a weakly tangled trajectory that encodes the desired figure-eight output in one projection and is a circle in another projection (Figure 7A, right subpanel). This is a natural shape to introduce: a circle is the least-tangled rhythmic trajectory.

Hypothesis-Based Prediction of Neural Responses

The results above suggest a hypothesis: motor cortex may embed outgoing commands (which, if muscle-like, would be quite tangled) in a larger trajectory such that the full orbit is minimally tangled. Inspired by optimizations that predicted V1 responses (Olshausen and Field, 1996), we employed an optimization approach to predict the dominant patterns of motor cortex activity. Optimization found a predicted neural population response, $\widehat{X}$, that could be linearly decoded to produce the empirical muscle activity Z, yet was minimally tangled. Specifically:

$$\widehat{X} = \underset{X}{\operatorname{argmin}} \left(\| Z - ZX^{\dagger}X \|_F^2 + \lambda \sum_t Q_X(t) \right), \qquad \text{(Equation 2)}$$

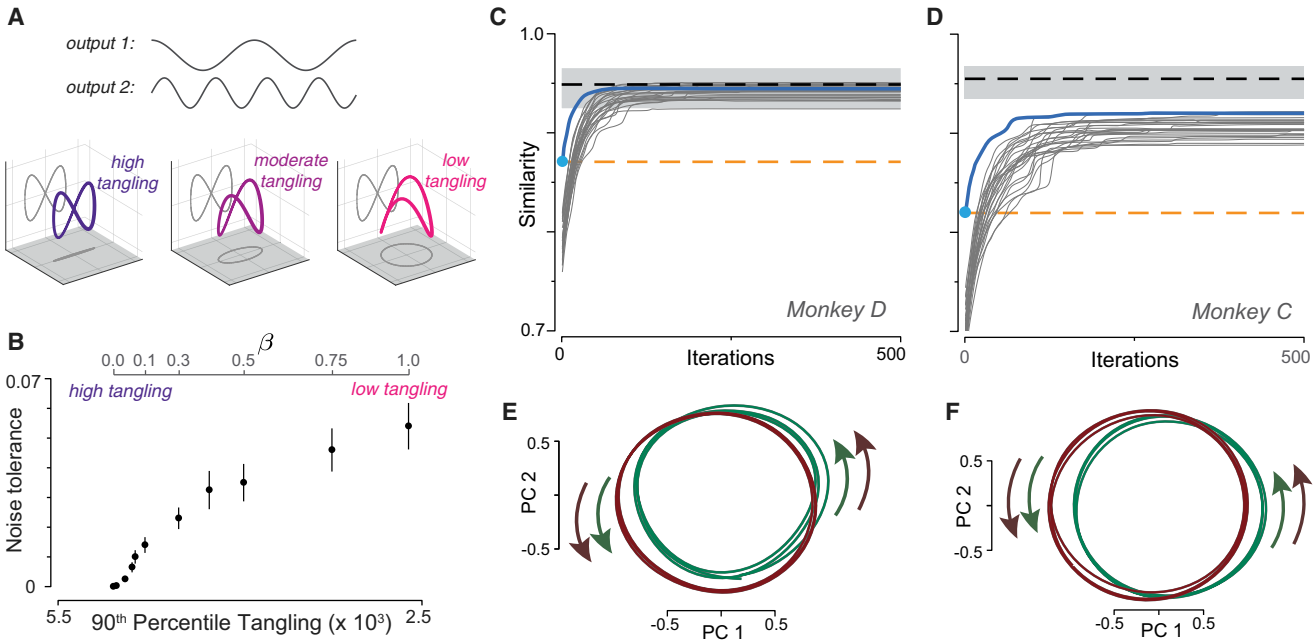

Figure 7. Low Trajectory Tangling Aids Noise Robustness and Can Be Leveraged to Predict the Motor Cortex Population Response

(A) Illustration of how an output can be embedded in a larger trajectory with varying degrees of tangling. Top traces: a hypothetical two-dimensional output [cos t; sin $2t$]. Plotted in state space, that output trajectory is a figure eight and contains a highly tangled central point. Adding a third dimension (βsin t) reduces tangling.

(B) Noise robustness of recurrent networks trained to follow the internal trajectory [cos t; sin $2t$; βsin t]. By varying β, we trained a set of networks that could all produce the same output but had varying degrees of trajectory tangling. Noise tolerance (mean and SEM across initializations) is plotted versus network trajectory tangling for each value of β.

(C) Similarity of the predicted and empirical motor cortex population responses (monkey D). Blue trace: prediction yielded by optimizing the cost function in Equation 2. Cyan dot indicates similarity at initialization; i.e., the similarity of empirical neural and muscle trajectories. This also provides a lower benchmark (orange dashed line). Gray traces: same as blue trace but with Gaussian noise added during initialization. Multiple initializations yielded a family of predictions. Black dashed line shows upper benchmark as described in the text, with a 95% confidence interval computed across random divisions of the population.

(D) Same but for monkey C.

(E) Projection of a representative predicted population response onto the top two PCs. Prediction based on EMG from monkey D. Green/red traces show trajectories for three cycles of forward/backward cycling.

(F) Same but for monkey C.

See also Figures S6 and S7.

where each column of the matrix Z describes the muscle population response for one time and condition. The first term of the cost function ensures that neural activity "encodes" muscle activity; $ZX^{\dagger}X$ is the optimal linear reconstruction of Z from X († indicates the pseudo-inverse; $\|\cdot\|_F$ indicates the Frobenius norm). This formulation should not be taken to imply that the true neural-to-muscle mapping is linear, merely that the predicted neural activity should yield a reasonable linear readout of muscle activity, consistent with empirical findings (Griffin et al., 2008; Morrow et al., 2009; Schieber and Rivlis, 2007). The second term of the cost function encourages low trajectory tangling. The predicted neural population response thus balances optimal encoding of muscle activity with minimal tangling.

We applied optimization using muscle data during three middle cycles of forward cycling and three middle cycles of backward cycling. Thus, we are attempting to simultaneously predict two "steady-state" neural trajectories. We used canonical correlation to assess similarity between predicted and actual neural responses. Canonical correlation finds linear transformations of two datasets such that they are maximally correlated. We employed a variant of canonical correlation that enforces orthonormal transformations. Unity similarity thus indicates two datasets are the same but for a rotation, isotropic scaling, or offset. We initialized optimization with $\hat{X}_{init} = Z$, corresponding to the baseline hypothesis that neural activity is a "pure" code for muscle activity. This yielded reasonably high initial similarity (Figures 7C and 7D, cyan dot) because muscle activity shares many basic features with neural activity (e.g., the same fundamental frequency).

During optimization, we insisted that the predicted neural population response, $\hat{X}$, have the same dimensionality as the muscle population response, Z (both were ten dimensional). Matching dimensionality is a conservative choice that aids interpretation. When optimization cannot add dimensions, some muscle-like features must be lost in order to gain features that reduce tangling. Similarity will therefore increase only if the features

gained during optimization are more realistic/prominent than the features that are lost.

Similarity between predicted and empirical populations increased with optimization (Figures 7C and 7D, blue), reaching a similarity roughly halfway between the "pure muscle encoding" hypothesis and perfect similarity. To provide a rough benchmark of good similarity, we computed the average similarity between two random halves of the empirical neural population (black dashed trace with 95% confidence intervals). Similarity approached this benchmark for both monkeys. To assess this result's consistency, we repeated optimization, each time initializing with the empirical patterns of muscle activity plus temporally smooth noise in each of the ten dimensions. Similarity to the data always increased during optimization (gray traces). This analysis also revealed that adding random structure decreases initial similarity (gray traces start below the blue trace). This underscores that similarity increased during optimization due to the introduction of structure matching that in the neural data, and not simply any arbitrary structure.

Each initialization resulted in a slightly different solution (the optimized $\hat{X}$). We were thus able to ask which solutions were common and whether the nature of those solutions explains the increased similarity with the empirical data. For all 200 solutions (100 per monkey), optimization produced near-circular trajectories. When comparing between forward and backward, two classes of solutions emerged. The less common (31/100 for monkey D and 13/100 for monkey C) involved dominant circular trajectories in planes that were nearly orthogonal (first principal angle >85°) for forward and backward. The most common (69/100 and 87/100 for monkey D and C) involved at least some overlap between these planes. In such cases, trajectories were almost always co-rotational (67/69 and 85/87 for monkey D and C) in the top two PCs. Two typical solutions are shown in Figures 7E and 7F. Co-rotations dominate because, when trajectories exist in a common subspace, tangling is lowest if they co-rotate (if they exist in orthogonal planes, co-rotation versus counter-rotation is not defined). Similar structure was seen for the empirical data: the planes that best captured neural trajectories during forward and backward cycling overlapped (principal angles were 72° and 61° for monkey D, and 73° and 40° for monkey C) and showed co-rotation in the top two PCs (as in Figures 4E and 4H). Thus, optimization based on Equation 2 not only increased quantitative similarity, it reproduced the dominant features of the neural data: nearly circular trajectories that exist in distinct but overlapping planes, and that co-rotate in the projection capturing the most variance.

Alternative Predictions

We performed a variety of optimizations corresponding to cost functions embodying other hypotheses (Figure S7). Optimizations that sought to reduce the norm of activity or to increase sparseness (standard forms of regularization) decreased similarity. Optimizing for local smoothness (one aspect of low tangling) increased similarity but not as effectively as optimizing for low tangling itself. Thus, similarity increased only when optimization reduced tangling and increased most when low tangling was directly optimized.

However, low tangling per se was not sufficient to increase similarity. We created simulated populations where the response of each unit was either the response of a muscle or the derivative of that response. This reflects the hypothesis that neurons might represent both muscle activity and the change in muscle activity (Evarts, 1968). By construction, these simulated populations had fairly low tangling (Figure S8A). Yet, they did not particularly resemble the neural population. Quantitatively, similarity increased modestly for monkey D (roughly half as much as when optimizing for low tangling directly) and decreased for monkey C. The dominant signals in these simulated populations did not show the same dominant circular structure seen in the neural data (Figure S8B). The mismatch can be understood by noting that differentiation increases the prevalence of high-frequency features. This does not lead to a match with the dominant circular structure at the fundamental frequency in the empirical data. In summary, optimizing directly for low tangling introduced features that were both particularly effective in reducing tangling and matched features in the data. Reducing tangling in a more "incidental" fashion did not produce these realistic features.

Signals Introduced by Optimization Yield Incidental Correlations

The optimization based on Equation 2 added structure that reduced tangling. That structure is unconnected to kinematics—of which optimization had no knowledge. Nevertheless, the predicted neural population response appeared to encode kinematics to a greater degree than would a pure code for muscle activity. We used linear regression to decode a set of kinematic parameters (horizontal and vertical position and velocity) from the activity of the muscle population. Fits were reasonable (R^2 = 0.86 and 0.88 for monkey D and C) but improved (R^2 = 0.97 and 0.94) when we instead decoded kinematics from the predicted neural population response. This performance was nearly identical to that observed when decoding kinematics from the empirical neural population (R^2 = 0.98 and 0.93). The ability to decode horizontal and vertical velocity might initially seem surprising: the dominant signals in the neural data co-rotated in the top two PCs—inconsistent with a velocity representation. However, the presence of more than two dimensions with sinusoidal structure ensured that velocity could be read out reasonably accurately.

Despite these excellent decodes, generalization performance was poor: generalization R^2 was near-zero (or even negative) when fitting kinematics for one direction and predicting for the other. This was true whether decoding was based on the predicted or empirical neural response. While poor generalization does not exclude the possibility that the empirical population encodes kinematic signals, we saw no direct evidence for this hypothesis. As noted above, we also rarely observed neurons whose firing rates resembled kinematic parameters.

Muscle-like Signals Are Embedded in Trajectories with Low Tangling

The success of optimization based on Equation 2 suggests a hypothesis: the dominant population-level signals in motor cortex may function to yield low tangling, with muscle-like signals encoded by relatively modest "ripples" in dimensions that point off the plane of dominant circular structure. A rough analogy would be a phonograph, where the direction encoding the temporally complex output is orthogonal to the dominant

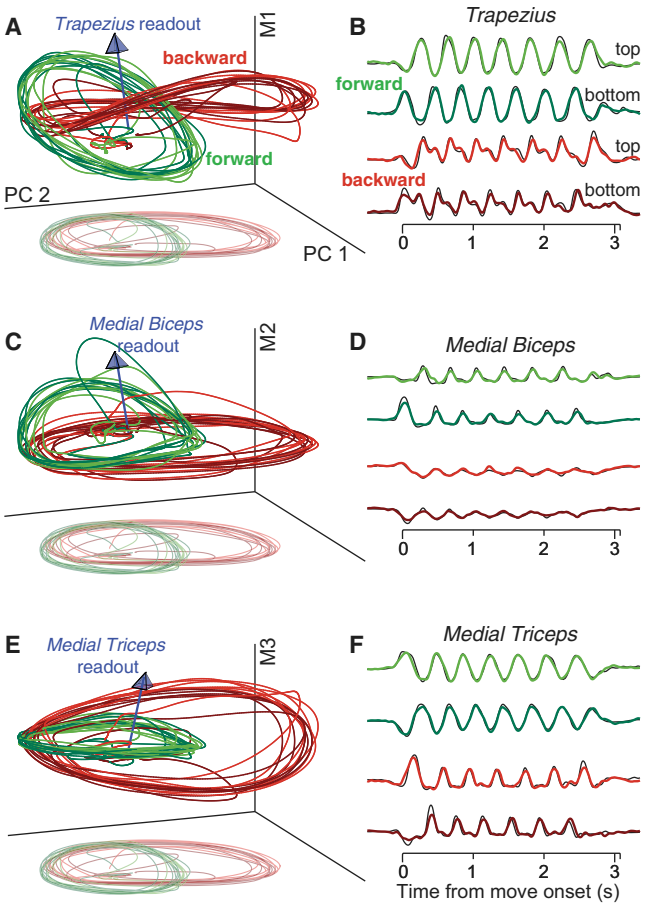

Figure 8. Muscle-like Signals Coexist with Signals that Contribute to Low Tangling

Data are for monkey D.

(A) Three-dimensional subspace capturing trajectories that can be linearly read out to decode trapezius activity. Blue arrow indicates the readout direction. "Shadow" traces at bottom show the projection onto just the first two PCs (perspective has been added). Trajectories are shown for four conditions: forward and backward seven-cycle movements, starting at the top and bottom (lighter and darker traces).

(B) Projections onto the readout direction. Thin black traces plot the empirical trapezius activity.

(C and D) Same as (A) and (B) but for medial biceps. Only the third (vertical) axis is different.

(E and F) Same but for medial triceps.

See also Figure S9.

motion of the record. Can such structure be viewed in the empirical data? We projected the neural population response onto triplets of dimensions (Figure 8). The first and second dimensions were always the first two PCs. The third was based on the readout direction of a particular muscle, defined by the set of weights found via linear regression. The third dimension was then the vector that was both orthogonal to the first two PCs and allowed the three dimensions to span the readout direction. Consider first a triplet of dimensions spanning the trapezius readout direction (Figure 8A). Trajectories trace out circular paths in the top PCs. Ripples in a third dimension provide the fine temporal structure that matches trapezius activity Figure 8B). The overall trajectory thus has the joint properties of encoding trapezius activity while exhibiting low tangling. Similar structure was observed for other muscles (Figures 8C and 8E; data for monkey C are shown in Figure S9).

The dimensions that encode muscle activity captured only modest variance. In the examples in Figure 8, each muscle-readout dimension captured ~10% as much variance as the average of the top two PCs. The vertical dimensions in Figures 8A, 8C, and 8E are thus shown on an expanded scale for visualization. A similar structure was present for the network model in Figure 5C and also for the predicted population responses in Figures 7E and 7F: the activity of each encoded muscle constituted a set of ripples upon dominant circular structure that yielded low tangling.

In addition to dimensions from which muscle-like signals can be read out, there exist other dimensions (not visible in Figure 8) that provide separation between neural trajectories during forward and backward cycling. Low tangling may require such separation—otherwise forward and backward trajectories would need to encode very different patterns of muscle activity despite following similar paths. Indeed, forward and backward neural trajectories were on average much better separated than the corresponding muscle trajectories (Figure S10). This difference in separation was large but not as profound as the difference in tangling. Thus, low neural-trajectory tangling (relative to muscle-trajectory tangling) results from a variety of factors: more circular trajectories, increased separation between forward and backward trajectories, and greater alignment of flow-fields (e.g., co-rotation in the dominant dimensions).

Tangling in Sulcal Motor Cortex

The results above support the hypothesis that population activity in motor cortex is less tangled than the outputs of that population. If so, tangling might be predicted to be moderately higher in sulcal motor cortex, where some neurons (cortico-motoneurons) make mono-synaptic connections onto spinal motoneurons (Rathelot and Strick, 2009), and signals related to outgoing muscle-like commands might thus be enriched. This is worth investigating both as an additional test of the central hypothesis, and because our measurements of muscle activity are only a proxy for the output of motor cortex. Ideally, we would be able to compute tangling for a subpopulation of identified cortico-motoneurons. In the absence of such recordings, we considered the subpopulation of sulcal recordings as a whole and compare with a subpopulation from the most anterior region from which we recorded: the aspect of dorsal premotor cortex contiguous with surface primary motor cortex. Cortico-motoneurons are largely absent from this anterior region (Rathelot and Strick, 2006). The subpopulation of sulcal neurons did indeed show significantly higher tangling during both cycling and reaching (Figure S11).

DISCUSSION

Are the Dominant Signals in Motor Cortex Representational or Computational?

We found that the dominant signals in motor cortex were not muscle-like. This result echoes findings during reaching, where

aspects of neural responses depart from expectations under a muscle-encoding framework (Evarts, 1968; Heming et al., 2016; Kakei et al., 1999; Moran and Schwartz, 1999b; Scott, 2008; Scott and Kalaska, 1997; Todorov, 2000). The dominance of non-muscle-like signals is more patent during cycling; non-muscle-like signals are apparent simply via inspection of projections onto the top PCs.

A traditional explanation for non-muscle-like signals is that they represent higher-level movement parameters. The present results are inconsistent with the most common proposal: a representation of direction or velocity. Under that proposal, trajectories should have been co-planar and counter-rotated between forward and backward cycling. We also found that single-neuron responses rarely resembled velocity profiles. Our data do not rule out the possibility that neural activity encodes a yet-to-be-determined set of kinematic parameters (perhaps in addition to muscle-like signals). However, our results urge caution when considering such hypotheses. For example, reducing tangling via optimization increases the degree to which activity appears (incorrectly) to represent kinematic parameters. More broadly, it may often be possible post hoc to select kinematic parameters that resemble the dominant neural signals, but this may generalize poorly across tasks. As one example, a representation of horizontal position and velocity would produce ellipses that co-rotate during forward/backward cycling. However, this "horizontal kinematics" hypothesis would require a high relative position sensitivity to ensure a circular trajectory. A high position sensitivity is inconsistent with observations during reaching, where correlations are strongest with reach velocity and direction (Ashe and Georgopoulos, 1994). In summary, in this study as in others, there will always be correlations that are incidental rather than fundamental (Churchland and Shenoy, 2007; Fetz, 1992; Mussa-Ivaldi, 1988; Reimer and Hatsopoulos, 2009; Todorov, 2000). While it remains possible that kinematic parameters are represented, we saw no compelling evidence for this idea. The dominant signals were already naturally explained by the hypothesis that tangling should be minimized. Furthermore, the observation of low tangling generalized well across tasks.

Our results thus suggest that the dominant signals in cortex may play a computational rather than a representational function. Specifically, the dominant signals may fall partly or largely in the null space of communication with downstream structures yet may be critical for ensuring reliable generation of the commands that are communicated. Put differently, motor cortex is part of a larger dynamical system (spanning many areas, including the spinal cord, and incorporating sensory feedback) that culminates in the generation of muscle commands. Such a system as a whole is almost certain to contain non-output signals. It does logically follow that motor cortex itself must show either non-output signals or low tangling; motor cortex could be downstream of the relevant dynamics or reflect only a small part of the overall network state. Yet empirically, motor cortex displayed very low tangling.

Differences and Commonalities across Tasks

During both cycling and reaching (Churchland et al., 2012), neural trajectories follow circular paths that rotate in a concordant direction, a feature not seen in the muscle population during either task. This shared feature may reflect the combination of two facts. First, a circle is the least-tangled rhythmic trajectory. Second, muscle activity during both tasks involves rhythmic aspects. This is trivially true during cycling. It is more subtly true during reaching, where multiphasic patterns of muscle activity are readily constructed from a quasi-oscillatory basis (Churchland and Cunningham, 2014; Churchland et al., 2012). Rotational trajectories are thus a natural way of encoding muscle activity while maintaining low tangling. This interpretation agrees with the recent finding that a network model, trained to produce muscle activity during reaching, produced rotational neural trajectories (Sussillo et al., 2015). This occurred only if the network was regularized to encourage smooth dynamics, a regularization that would implicitly encourage low tangling.

Still, we stress that rotational structure per se is unlikely to be the fundamental principle shared across tasks. There are many ways of adding structure that can reduce tangling. Even if certain motifs are common, the optimal way to reduce tangling will be task dependent. Thus, we propose that the deeper connection across tasks will not be a specific form of dynamics, but dynamics that yield low tangling.

We also note that different tasks may involve motor cortex sending different classes of output commands. For some tasks, the details of muscle activity may be largely determined by spinal circuitry, while other tasks (especially learned or dexterous tasks) may require more direct control of the musculature. The latter is potentially true during cycling, and some of our analyses thus assumed a roughly linear relationship between neural and muscle activity. However, the hypothesized computational principle—embed outgoing commands in structure that minimizes tangling—would apply even if commands were only somewhat muscle-like (e.g., if they were transformed considerably by the spinal cord). Indeed, it would apply even if descending commands are high-level, as may have been the case in mice during locomotion.

Tangling across Areas

Trajectory tangling was very low for motor cortex, considerably higher for S1, and higher still for the muscles. Tangling was also high for V1. The degree of tangling may depend on how fully activity in that area reflects the relevant global network and feedback dynamics. Motor cortex may show particularly low tangling because it processes many relevant sources of information. It is not only a major output of the primate motor system but responds robustly and rapidly to sensory inputs (Herter et al., 2009) and lies at the nexus of cerebellar and basal-ganglia feedback loops (Middleton and Strick, 2000). Other areas, even those that participate in the same task, may or may not exhibit low tangling depending on how fully they reflect the overall network state. In particular, S1 responses are likely dominated by sensory feedback and may very incompletely reflect the broader dynamics of motor control. Even within motor cortex, tangling was modestly higher within the sulcus, where activity may be more dominated by output commands. Although V1 presumably does exhibit some dynamics, activity is likely dominated by visual inputs, which can produce high tangling. These comparisons echo our recent finding that population structure can be fundamentally different depending on whether an area is hypothesized to primarily reflect population dynamics versus external variables (Seely et al., 2016).

Differences between areas raise the question of whether tangling might sometimes differ within a population. Might the motor system, over the course of learning or development, adopt network trajectories that are increasingly less tangled? When a new skill is learned, is performance better if subjects achieve lower tangling? Are pathological conditions associated with increased tangling? Such questions illustrate that many aspects of motor cortex activity may be best understood not in terms of representations of external parameters, but in terms of the computational strategies that allow outputs to be accurately and reliably generated.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Main Experimental Datasets
  - Reaching Datasets
  - Visual Cortex Datasets
  - Mouse Datasets
- METHOD DETAILS
  - Task
  - Neural Recordings during Cycling
  - EMG Recordings
  - Trial Alignment and Averaging
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Preprocessing and PCA
  - Regression
  - Tangling
  - Computational Motivation for the Tangling Measure
  - Standard Recurrent Neural Networks
  - Trajectory-Constrained Neural Networks
  - Predicting Neural Population Activity
  - Similarity between Empirical and Predicted Data
  - Predictions via Alternate Cost Functions

SUPPLEMENTAL INFORMATION

Supplemental Information includes eleven figures and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.01.004.

ACKNOWLEDGMENTS

We thank C. Hussar for task development and Y. Pavlova for animal care. Support was provided by the Grossman Center for the Statistics of Mind, Burroughs Wellcome Fund (M.M.C.), Searle Scholars Program (M.M.C.), Sloan Foundation (M.M.C. and J.P.C.), Simons Foundation (M.M.C., J.P.C., L.F.A., T.M.J., and A.K.), McKnight Foundation (M.M.C. and J.P.C.), Helen Hay Whitney Foundation (A.M.), NIH Director's New Innovator Award DP2 NS083037 (M.M.C.), NIH NS033245 (T.M.J.), NIH EY016774 (A.K.), NIH CRCNS R01NS100066 (M.M.C. and J.P.C.), NIH 1U19NS104649 (M.M.C., L.F.A., and T.M.J.), NIH R01MH93338 (L.F.A.), NIH F32NS092350 (A.H.L.), NIH 5T32NS064929 (A.A.R.), National Science Foundation (N.J.M., J.S.S., and S.R.B.), Kavli Foundation (M.M.C. and T.M.J.), Klingenstein Foundation (M.M.C.), Project ALS (T.M.J.), Mathers Foundation (T.M.J.), and the Howard Hughes Medical Institute (T.M.J.).

AUTHOR CONTRIBUTIONS

M.M.C. conceived the experiments; A.A.R., B.M.L., and S.M.P. collected main datasets. N.J.M. contributed some EMG recordings; A.A.R. and M.M.C. designed data analyses, aided by J.P.C. and L.F.A.; A.A.R. performed analyses. S.R.B. and J.P.C. performed predictive optimizations. S.R.B. and J.S.S. trained network models, supervised by L.F.A.; A.H.L. collected reaching datasets; A.M. collected rodent dataset, with T.M.J. and A.M.; A.K. collected V1 dataset; A.A.R. and M.M.C. wrote the paper. All authors contributed to editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 29, 2017
Revised: October 24, 2017
Accepted: December 31, 2017
Published: February 1, 2018

REFERENCES

Aflalo, T.N., and Graziano, M.S.A. (2007). Relationship between unconstrained arm movements and single-neuron firing in the macaque motor cortex. J. Neurosci. *27*, 2760–2780.

Ajemian, R., Green, A., Bullock, D., Sergio, L., Kalaska, J., and Grossberg, S. (2008). Assessing the function of motor cortex: single-neuron models of how neural response is modulated by limb biomechanics. Neuron *58*, 414–428.

Ashe, J., and Georgopoulos, A.P. (1994). Movement parameters and neural activity in motor cortex and area 5. Cereb. Cortex *4*, 590–600.

Burnod, Y., Grandguillaume, P., Otto, I., Ferraina, S., Johnson, P.B., and Caminiti, R. (1992). Visuomotor transformations underlying arm movements toward visual targets: a neural network model of cerebral cortical operations. J. Neurosci. *12*, 1435–1453.

Cheney, P.D., and Fetz, E.E. (1980). Functional classes of primate corticomotoneuronal cells and their relation to active force. J. Neurophysiol. *44*, 773–791.

Churchland, M.M., and Cunningham, J.P. (2014). A dynamical basis set for generating reaches. Cold Spring Harb. Symp. Quant. Biol. *79*, 67–80.

Churchland, M.M., and Shenoy, K.V. (2007). Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. J. Neurophysiol. *97*, 4235–4257.

Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56.

Cunningham, J.P., and Ghahramani, Z. (2015). Linear dimensionality reduction: survey, insights, and generalizations. J. Mach. Learn. Res. *16*, 2859–2900.

Druckmann, S., and Chklovskii, D.B. (2012). Neuronal circuits underlying persistent representations despite time varying activity. Curr. Biol. *22*, 2095–2103.

Elsayed, G.F., Lara, A.H., Kaufman, M.T., Churchland, M.M., and Cunningham, J.P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nat. Commun. *7*, 13239.

Evarts, E.V. (1968). Relation of pyramidal tract activity to force exerted during voluntary movement. J. Neurophysiol. *31*, 14–27.

Fetz, E.E. (1992). Are movement parameters recognizably coded in the activity of single neurons? Behav. Brain Sci. *15*, 679–690.

Fitzsimmons, N.A., Lebedev, M.A., Peikon, I.D., and Nicolelis, M.A. (2009). Extracting kinematic parameters for monkey bipedal walking from cortical neuronal ensemble activity. Front. Integr. Nuerosci. *3*, 3.

Foster, J.D., Nuyujukian, P., Freifeld, O., Gao, H., Walker, R., I Ryu, S., H Meng, T., Murmann, B., J Black, M., and Shenoy, K.V. (2014). A freely-moving monkey treadmill model. J. Neural Eng. *11*. Published online July 4, 2014. https://doi.org/10.1088/1741-2560/11/4/046020.

Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., and Massey, J.T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J. Neurosci. *2*, 1527–1537.

Georgopoulos, A.P., Schwartz, A.B., and Kettner, R.E. (1986). Neuronal population coding of movement direction. Science *233*, 1416–1419.

Georgopoulos, A.P., Naselaris, T., Merchant, H., and Amirikian, B. (2007). Reply to kurtzer and herter. J. Neurophysiol. *97*, 4391–4392.

Griffin, D.M., Hudson, H.M., Belhaj-Saïf, A., McKiernan, B.J., and Cheney, P.D. (2008). Do corticomotoneuronal cells predict target muscle EMG activity? J. Neurophysiol. *99*, 1169–1986.

Hall, T.M., de Carvalho, F., and Jackson, A. (2014). A common structure underlies low-frequency cortical dynamics in movement, sleep, and sedation. Neuron *83*, 1185–1199.

Hart, C.B., and Giszter, S.F. (2010). A neural basis for motor primitives in the spinal cord. J. Neurosci. *30*, 1322–1336.

Hatsopoulos, N.G., Xu, Q., and Amit, Y. (2007). Encoding of movement fragments in the motor cortex. J. Neurosci. *27*, 5105–5114.

Heming, E.A., Lillicrap, T.P., Omrani, M., Herter, T.M., Pruszynski, J.A., and Scott, S.H. (2016). Primary motor cortex neurons classified in a postural task predict muscle activation patterns in a reaching task. J. Neurophysiol. *115*, 2021–2032.

Hennequin, G., Vogels, T.P., and Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. Neuron *82*, 1394–1406.

Herter, T.M., Korbel, T., and Scott, S.H. (2009). Comparison of neural responses in primary motor cortex to transient and continuous loads during posture. J. Neurophysiol. *101*, 150–163.

Kakei, S., Hoffman, D.S., and Strick, P.L. (1999). Muscle and movement representations in the primary motor cortex. Science *285*, 2136–2139.

Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2014). Cortical activity in the null space: Permitting preparation without movement. Nat. Neurosci. *17*, 440–448.

Kaufman, M.T., Seely, J.S., Sussillo, D., Ryu, S.I., Shenoy, K.V., and Churchland, M.M. (2016). The largest response component in the motor cortex reflects movement timing but not movement type. eNeuro *3*. Published online August 30, 2016. https://doi.org/10.1523/ENEURO.0085-16.2016.

Li, N., Daie, K., Svoboda, K., and Druckmann, S. (2016). Robust neuronal dynamics in premotor cortex during motor planning. Nature *532*, 459–464.

Lillicrap, T.P., and Scott, S.H. (2013). Preference distributions of primary motor cortex neurons reflect control solutions optimized for limb biomechanics. Neuron *77*, 168–179.

Maier, M.A., Shupe, L.E., and Fetz, E.E. (2005). Dynamic neural network models of the premotoneuronal circuitry controlling wrist movements in primates. J. Comput. Neurosci. *19*, 125–146.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84.

Michaels, J.A., Dann, B., and Scherberger, H. (2016). Neural population dynamics during reaching are better explained by a dynamical system than representational tuning. PLOS Comput. Biol. *12*, e1005175.

Middleton, F.A., and Strick, P.L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res. Brain Res. Rev. *31*, 236–250.

Miri, A., Warriner, C.L., Seely, J.S., Elsayed, G.F., Cunningham, J.P., Churchland, M.M., and Jessell, T.M. (2017). Behaviorally selective engagement of short-latency effector pathways by motor cortex. Neuron *95*, 683–696.e11.

Moran, D.W., and Schwartz, A.B. (1999a). Motor cortical activity during drawing movements: population representation during spiral tracing. J. Neurophysiol. *82*, 2693–2704.

Moran, D.W., and Schwartz, A.B. (1999b). Motor cortical representation of speed and direction during reaching. J. Neurophysiol. *82*, 2676–2692.

Moran, D.W., and Schwartz, A.B. (2000). One motor cortex, two different views. Nat. Neurosci. *3*, 963, author reply 963–965.

Morrow, M.M., Pohlmeyer, E.A., and Miller, L.E. (2009). Control of muscle synergies by cortical ensembles. Adv. Exp. Med. Biol. *629*, 179–199.

Mussa-Ivaldi, F.A. (1988). Do neurons in the motor cortex encode movement direction? An alternative hypothesis. Neurosci. Lett. *91*, 106–111.

Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature *381*, 607–609.

Rathelot, J.A., and Strick, P.L. (2006). Muscle representation in the macaque motor cortex: an anatomical perspective. Proc. Natl. Acad. Sci. USA *103*, 8257–8262.

Rathelot, J.A., and Strick, P.L. (2009). Subdivisions of primary motor cortex based on cortico-motoneuronal cells. Proc. Natl. Acad. Sci. USA *106*, 918–923.

Reimer, J., and Hatsopoulos, N.G. (2009). The problem of parametric neural coding in the motor system. Adv. Exp. Med. Biol. *629*, 243–259.

Rokni, U., and Sompolinsky, H. (2012). How the brain generates movement. Neural Comput. *24*, 289–331.

Sanger, T.D. (1994). Theoretical considerations for the analysis of population coding in motor cortex. Neural Comput. *6*, 29–37.

Schieber, M.H., and Rivlis, G. (2007). Partial reconstruction of muscle activity from a pruned network of diverse motor cortex neurons. J. Neurophysiol. *97*, 70–82.

Schwartz, A.B. (1994). Direct cortical representation of drawing. Science *265*, 540–542.

Schwartz, A.B. (2007). Useful signals from motor cortex. J. Physiol. *579*, 581–601.

Schwartz, A.B., Moran, D.W., and Reina, G.A. (2004). Differential representation of perception and action in the frontal cortex. Science *303*, 380–383.

Scott, S.H. (2008). Inconvenient truths about neural processing in primary motor cortex. J. Physiol. *586*, 1217–1224.

Scott, S.H., and Kalaska, J.F. (1997). Reaching movements with similar hand paths but different arm orientations. I. Activity of individual cells in motor cortex. J. Neurophysiol. *77*, 826–852.

Seely, J.S., Kaufman, M.T., Ryu, S.I., Shenoy, K.V., Cunningham, J.P., and Churchland, M.M. (2016). Tensor analysis reveals distinct population structure that parallels the different computational roles of areas M1 and V1. PLoS Comput. Biol. *12*, e1005164.

Sergio, L.E., Hamel-Pâquet, C., and Kalaska, J.F. (2005). Motor cortex neural correlates of output kinematics and kinetics during isometric-force and arm-reaching tasks. J. Neurophysiol. *94*, 2353–2378.

Shalit, U., Zinger, N., Joshua, M., and Prut, Y. (2012). Descending systems translate transient cortical commands into a sustained muscle activation signal. Cereb. Cortex *22*, 1904–1914.

Shenoy, K.V., Sahani, M., and Churchland, M.M. (2013). Cortical control of arm movements: a dynamical systems perspective. Annu. Rev. Neurosci. *36*, 337–359.

Sussillo, D., and Abbott, L.F. (2009). Generating coherent patterns of activity from chaotic neural networks. Neuron *63*, 544–557.

Sussillo, D., and Barak, O. (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. Neural Comput. *25*, 626–649.

Sussillo, D., Churchland, M.M., Kaufman, M.T., and Shenoy, K.V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. Nat. Neurosci. *18*, 1025–1033.

Todorov, E. (2000). Direct cortical control of muscle activation in voluntary arm movements: a model. Nat. Neurosci. *3*, 391–398.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Rhesus macaque (Mucacca mulatta)	Davis National Primate Center	N/A
Software and Algorithms		
MATLAB	MathWorks	https://www.mathworks.com/products/matlab.html
Simulink	MathWorks	https://www.mathworks.com/products/simulink-real-time.html
Unity Engine	Unity Technologies	https://unity3d.com/
Python	Python Software Foundation	https://www.python.org/
Other		
Speedgoat Real-time Target Machine	Speedgoat	https://www.speedgoat.com/products-services/real-time-target-machines/performance
Polaris Eye Tracking System	Northern Digital	https://www.ndigital.com/medical/products/polaris-family/
Cerebus system	Blackrock Microsystems	http://blackrockmicro.com/neuroscience-research-products/neural-data-acquisition-systems/cerebus-daq-system/
Utah array	Blackrock Microsystems	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Mark M. Churchland (mc3502@columbia.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Main Experimental Datasets

Subjects were two adult male rhesus macaques (monkeys D and C). Animal protocols were approved by the Columbia University Institutional Animal Care and Use Committee. Experiments were controlled and data collected under computer control (Speedgoat Real-time Target Machine). During experiments, monkeys sat in a customized chair with the head restrained via a surgical implant. Stimuli were displayed on a monitor in front of the monkey. A tube dispensed juice rewards. The left arm was loosely restrained using a tube and a cloth sling. With their right arm, monkeys manipulated a pedal-like device. The device consisted of a cylindrical rotating grip (the pedal), attached to a crank-arm, which rotated upon a main axel. That axel was connected to a motor and a rotary encoder that reported angular position with 1/8000 cycle precision. In real time, information about angular position and its derivatives was used to provide virtual mass and viscosity, with the desired forces delivered by the motor. The delay between encoder measurement and force production was 1 ms.

Horizontal and vertical hand position were computed based on angular position and the length of the crank-arm (64 mm). To minimize extraneous movement, the right wrist rested in a brace attached to the hand pedal. The motion of the pedal was thus almost entirely driven by the shoulder and elbow, with the wrist moving only slightly to maintain a comfortable posture. Wrist movements were monitored via two reflective spheres attached to the brace, which were tracked optically (Polaris system; Northern Digital, Waterloo, Ontario, Canada) and used to calculate wrist angle. The small wrist movements were highly stereotyped across cycles. Visual monitoring (via infrared camera) confirmed the same was true of the arm as a whole (e.g., the lateral position of the elbow was quite stereotyped across revolutions). Eye position and pupil dilation were monitored but are not analyzed here.

Reaching Datasets

Recordings from primate motor cortex during reaching have been described and analyzed previously (Elsayed et al., 2016). Briefly, two male rhesus monkeys (A and B) performed center-out reaches in eight target directions on a fronto-parallel screen. This task employed three 'contexts' in which reach initiation was prompted by different cues. That manipulation was incidental to the present analysis: we analyzed only movement-related responses, which were empirically very similar across the three contexts. We therefore simply computed the trial-averaged time-varying firing rate (smoothed with a 20 ms SD. Gaussian) across all reaches for each of the eight directions. Trials were aligned to movement onset and we analyzed the period from 100 ms before movement onset until 100 ms after the average time of movement offset. Neural populations included 101 and 129 neurons (monkey A and B) recorded from the arm region of motor cortex (including sulcal and surface primary motor cortex and the adjacent aspect of dorsal premotor cortex).

During this same task, activity was recorded from the muscles of the upper arm (*deltoid, trapezius, biceps, brachialis, pectoralis, latissimus dorsi* muscles) using the same procedures described above (13 and 10 recordings for monkey A and B; smoothed with a 20 ms SD. Gaussian). The median number of analyzed trials per direction was 48 (monkey A) and 60 (monkey B).

Visual Cortex Datasets

Data from primate V1 were recorded using natural-movie stimuli from an anaesthetized adult monkey (*Macaca fascicularis*) implanted with a 96-electrode silicon 'Utah' array (Blackrock Microsystems, Salt Lake City, UT) in left-hemisphere V1 as previously described (Seely et al., 2016). These data were recorded in the laboratory of Adam Kohn. Procedures were approved by the Animal Care and Use Committees at Albert Einstein College of Medicine (protocol #20150303). The left eye was covered. Receptive field centers (2–4 degrees eccentric) were determined via brief presentations of small drifting gratings. Stimuli, which spanned the receptive fields, were 48 natural movie clips (selected from YouTube) with 50 repeats each. The frame rate was ~95 Hz. Each stimulus lasted 2.63 s (100 movie frames followed by 150 blank frames). Spikes from the array were sorted offline using MKsort (available at https://github.com/ripple-neuro/mksort/). A total of 108 single units and stable multi-unit isolations were included. It is unclear how anesthesia might affect trajectory tangling of this neural population. However, responses to stimuli were robust and only stimulus-evoked aspects of the responses were analyzed.

Mouse Datasets

Data from mouse motor cortex have been described and analyzed previously (Miri et al., 2017). Briefly, three head-fixed mice performed a task that included both a reach-to-grasp sub-task and natural treadmill walking (10 cm/s), performed in separate blocks. Multiple neurons / muscles were recorded simultaneously, but were also accumulated across days to allow analysis of larger populations. The populations for each mouse were analyzed separately. Neural recordings were made with independently movable tetrode micro-drives, lowered over the course of two weeks to primarily target layer 5. A total of 890 well-isolated units from three animals were recorded across 11 behavioral sessions. Muscle activity from the forelimb was recorded from electrodes chronically implanted in the *trapezius*, *pectoralis*, *biceps*, *triceps*, *extensor digitorum communis*, and *palmaris longus*. For two mice, recordings were made from all six of these muscles. For one mouse, recordings could only be made from four. Each muscle was recorded across eleven sessions. PCA thus extracted the top EMG signals across 66 total records for two mice and 44 for the other. Spike-trains and muscle activity were smoothed with a Gaussian filter (20 ms SD) and averaged across trials.

METHOD DETAILS

Task

The monitor displayed a virtual landscape, generated by the Unity engine (Unity Technologies, San Francisco). Surface texture and landmarks to each side provided visual cues regarding movement through the landscape. Movement was along a linear 'track'. One rotation of the pedal produced one arbitrary unit of movement. Targets on the landscape surface indicated where the monkey should stop for juice reward.

Each trial of the task began with the appearance of an initial target. To begin the trial, the monkey had to cycle to and to acquire the initial target (i.e., stop on it and remain stationary) within 5 s. Acquisition of the initial target yielded a small reward. After a 1000 ms hold period, the final target appeared at a prescribed distance. Following a randomized (500-1000 ms) delay period, a go-cue (brightening of the final target) was given. The monkey then had to cycle to acquire the final target. After remaining stationary in the final target for 1500 ms, the monkey received a large reward.

Successfully completing a trial necessitated satisfying a variety of constraints. Cycling had to begin between within 650 ms after the go cue. Once cycling began, the final target had to be reached within a distance-dependent time limit. The trial was aborted if this time elapsed (< 0.01% of trials for both monkeys), or if cycling speed dropped below a threshold before entering the final target (~1.5% of trials in monkey D and ~1.7% in monkey C). The trial was also aborted if the monkey moved past the final target (~1.5% / 0.6% of trials), or if the monkey acquired the final target and then moved while waiting for the reward (~0.6% / 0.3%). These constraints, combined with the monkeys' natural desire to receive reward quickly, produced movements that were both brisk and quite consistent across trials. The primary difference in behavior across trials was modest variation in overall movement duration (as illustrated in Figure 1). In rare cases, behavior on a successful trial differed notably from typical behavior for that condition. Such trials were removed prior to analysis.

The task included 20 conditions distinguishable by final target distance (half-, one-, two-, four-, and seven-cycles), initial starting position (top or bottom of the cycle), and cycling direction. Salient visual cues (landscape color) indicated whether cycling must be 'forward' (the hand moved away from the body at the top of the cycle) or 'backward' (the hand moved toward from the body at the top of the cycle) to produce forward virtual progress. Trials were blocked into forward and backward cycling. Other trials types were interleaved using a block-randomized design. We collected a median of 15 trials / condition for both monkeys

Neural Recordings during Cycling

After initial training, we performed a sterile surgery during which monkeys were implanted with a head restraint and recording cylinders. Cylinders (Crist Instruments, Hagerstown, MD) were placed surface normal to the cortex, centered over the border between

caudal PMd and primary motor cortex, located according to a previous magnetic resonance imaging scan. The skull within the cylinder was left intact and covered with a thin layer of dental acrylic. Electrodes were introduced through small (3.5 mm diameter) burr holes drilled by hand through the acrylic and skull, under ketamine / xylazine anesthesia. Neural recordings were made using conventional single electrodes (Frederick Haer Company, Bowdoinham, ME) driven by a hydraulic microdrive (David Kopf Instruments, Tujunga, CA).

Sequential recording with conventional electrodes (as opposed to simultaneous recording with an array) allowed us to acquire recordings from a broader range of sites, including sulcal sites inaccessible to many array techniques. Recording locations were guided via microstimulation, light touch, and muscle palpation protocols to confirm the trademark properties of each region. For motor cortex, recordings were made from primary motor cortex (both surface and sulcal) and the adjacent (caudal) aspect of dorsal premotor cortex. For most analyses, these recordings are analyzed together as a single motor cortex population (although see Figure S11). Motor cortex recordings were restricted to regions where microstimulation elicited responses in shoulder, upper arm, chest, and forearm. For one monkey, we also recorded from area 3a (proprioceptive primary motor cortex). These recordings (44 neurons) were made from the deeper aspects of the posterior bank of the central sulcus, where microstimulation did not produce movement.

Neural signals were amplified, filtered, and manually sorted using Blackrock Microsystems hardware (Digital Hub and 128-channel Neural Signal Processor). A total of 277 isolations were made across the two monkeys. Nearly all neurons that could be isolated in motor cortex were responsive during cycling. A modest number (21) of isolations were discarded due to low signal-to-noise ratios or insufficient trial counts. No further selection criteria were applied. On each trial, the spikes of the recorded neuron were filtered with a Gaussian (25 ms standard deviation; SD) to produce an estimate of firing rate versus time. These were then averaged across trials as described below.

EMG Recordings

Intra-muscular EMG was recorded from the major muscles of the arm, shoulder, and chest using percutaneous pairs of hook-wire electrodes (30mm x 27 gauge, Natus Neurology) inserted ~1 cm into the belly of the muscle for the duration of single recording sessions. Electrode voltages were amplified, bandpass filtered (10-500 Hz) and digitized at 1000 Hz. To ensure that recordings were of high quality, signals were visualized on an oscilloscope throughout the duration of the recording session. Recordings were aborted if they contained significant movement artifact or weak signal. That muscle was then re-recorded later. Offline, EMG records were high-pass filtered at 40 Hz and rectified. Finally, EMG records were smoothed with a Gaussian (25 ms SD, same as neural data) and trial averaged (see below). Recordings were made from the following muscles: the three heads of the *deltoid*, the two heads of the *biceps brachii*, the three heads of the *triceps brachii*, *trapezius*, *latissimus dorsi, pectoralis, brachioradialis, extensor carpi ulnaris, extensor carpi radialis, flexor carpi ulnaris, flexor carpi radialis*, and *pronator*. Recordings were made from 1-8 muscles at a time, on separate days from neural recordings. We often made multiple recordings for a given muscle, especially those that we have previously noted can display responses that vary with recording location (e.g., the *deltoid*).

Trial Alignment and Averaging

To preserve response features, it was important to compute the average firing rate across trials with nearly identical behavior. This was achieved by 1) training to a high level of stereotyped behavior, 2) discarding rare aberrant trials, and 3) adaptive alignment of individual trials prior to averaging. Because of the temporally extended nature of cycling movements, standard alignment procedures (e.g., locking to movement onset) often misalign responses later in the movement. For example, a seven-cycle movement lasted ~3500 ms. By the last cycle, a trial 5% faster than normal and a trial 5% slower than normal would thus be misaligned by 350 ms, or over half a cycle.

To ensure response features were not lost to misalignment, we developed a technique to adaptively align trials within a condition. First, trials were aligned on movement onset. Individual trials were then scaled so that all trials had the same duration (set to be the median duration across trials). Because monkeys usually cycled at a consistent speed (within a given condition) this brought trials largely into alignment: e.g., the top of each cycle occurred at nearly the same time for each trial. The adaptive alignment procedure was used to correct any remaining slight misalignments. The time-base for each trial was scaled so that the position trace on that trial closely matched the average position of all trials. This involved a slight non-uniform stretching, and resulted in the timing of all key moments – such as when the hand passed the top of the cycle – being nearly identical across trials. This ensured that high-frequency temporal response features (e.g., the small peak in Figure 1G) were not lost to averaging.

All variables of interest (firing rate, hand position, hand velocity, EMG, etc.) were computed on each trial before adaptive alignment. Thus, the above procedure never alters the magnitude of these variables, but simply aligns when those values occur across trials. The adaptive procedure was used once to align trials within a condition on a given recording session, and again to align data across recording sessions. This allowed, for example, comparison of neural and muscle responses on a matched time-base.

QUANTIFICATION AND STATISTICAL ANALYSIS

Preprocessing and PCA

Because PCA seeks to capture variance, it can be disproportionately influenced by differences in firing rate range (e.g., a neuron with a range of 100 spikes/s has 25 times the variance of a similar neuron with a range of 20 spikes/s). This concern is larger still for EMG,

where the scale is arbitrary and can differ greatly between recordings. The response of each neuron / muscle was thus normalized prior to application of PCA. EMG data were fully normalized: $response := response / range(response)$, where the range is taken across all recorded times and conditions. Neural data were 'soft' normalized: $response := response / (range(response) + 5)$. We standardly (Churchland et al., 2012; Seely et al., 2016) use soft normalization to balance the desire for PCA to explain the responses of all neurons with the desire that weak responses not contribute on an equal footing with robust responses. In practice, nearly all neurons had high firing rate ranges during cycling, making soft normalization nearly identical to full normalization.

Following preprocessing, neural data were formatted as a 'full-dimensional' matrix, X^{full}, of size $n \times t$, where n is the number of neurons and t indexes across all analyzed times and conditions. We similarly formatted muscle data as a matrix, Z^{full}, of size $m \times t$, where m is the number of muscles. Unless otherwise specified, analyzed times were from 100 ms before movement onset to 100 ms after movement offset, for all conditions. Because PCA operates on mean-centered data, we mean-centered X^{full} and Z^{full} so that every row had a mean value of zero.

PCA was used to find X, a reduced-dimensional version of X^{full} with the property that $X^{full} \approx VX$, where V are the PCs ('neural dimensions' upon which the data are projected). PCA was similarly used to find Z, the reduced-dimensional version of Z^{full}. For most analyses, we employed eight PCs, such that X and Z were of size $8 \times t$. Eight PCs captured 70% and 68% (monkey D and C) of the neural data variance, and 94% and 88% of the muscle data variance.

Regression

Decoding of muscle activity from neural activity was accomplished via a linear model: $Z^{full} = BX^{full}$. B was found using ridge regression. Performance was assessed using generalization R^2, using Leave-One-Out Cross Validation. Regularization strength was chosen to maximize Leave-One-Out Cross Validation performance, though in practice a broad range of regularization strengths provided similar performance. We also attempted to decode neural activity from muscle activity using the model $X^{full} = BZ^{full}$. Decoding neural activity from muscle activity was less successful than decoding muscle activity from neural activity. Although our neural recordings generally had very good signal-to-noise, we considered that poor decoding of neural activity from muscle activity (relative to decoding muscle activity from neural activity) could potentially result because neural responses tend to have higher sampling error than muscle responses. We therefore re-ran the regression above after de-noising the neural data by replacing each neuron's response with its reconstruction using the top thirty PCs. The same discrepancy was observed.

In a subsequent analysis, we decoded kinematic parameters from both predicted and empirical population activity. The predicted population response pertained only to the three middle cycles of seven-cycle movements. Thus, all decoding of kinematic parameters involved only those three cycles. Decoding employed ridge regression as described above. Regularization strength was chosen to improve generalization performance without overly sacrificing test performance. Kinematics were mean centered, and regressed against the ten dimensions of the predicted population response, or the projection of the empirical data onto the top ten PCs. Matching dimensionality ensured that it is appropriate to compare R^2 and generalization R^2 values when regressing against the predicted versus empirical population. Generalization performance was tested by fitting to data for one direction (e.g., forward cycling) and generalizing to the other (e.g., backward cycling).

Tangling

Tangling was computed as described in the results (Equation 1). The neural state, $\boldsymbol{x}_t$ was an 8×1 vector comprised of the t^{th} column of X, where X is of size $8 \times t$. Muscle tangling was computed analogously, based on Z. Essentially identical results were found if we used X^{full} and Z^{full} (Figure S2) but this was less computationally efficient and did not allow matched dimensionality between neurons and muscles. We computed the derivative of the state as $\dot{\boldsymbol{x}}_t = (\boldsymbol{x}_t - \boldsymbol{x}_{t-\Delta t})/\Delta t$, where Δt was 1 ms. When computing tangling, we employed the squared distance between derivatives, $\| \dot{\boldsymbol{x}}_t - \dot{\boldsymbol{x}}_{t'} \|^2$, because its magnitude more intuitively tracks the difference in trajectory direction. For example, if the angle between derivatives doubles from 90° to 180°, the norm grows by only 41%, but the squared norm is doubled. The constant ε was set to 0.1 times the average squared magnitude of $\boldsymbol{x}_t$ across all t. Results were essentially identical across an order of magnitude of values of ε.

Tangling estimates how non-smooth a flow-field would have to be to have produced the observed trajectories. While there are many potential measures one could use, tangling is simple to compute directly from the data, without any need to attempt to estimate the underlying flow-field. The simplicity of the tangling measure is desirable not only from a data analysis standpoint, but also from the standpoint of the optimizations in Figures 7 and S7. A more complicated measure would have resulted in a cost function that was difficult or impossible to minimize. The ability to compute tangling without fitting a flow-field is desirable because even with many conditions and temporally extended trajectories, the data leave many large 'gaps' in high-dimensional state space, making it difficult to fit an overall flow-field with any confidence. That said, one would still hope that tangling would correlate with how well the flow-field can be fit by a dynamical model with smoothness constraints (e.g., a linear model). This was indeed the case. Muscle trajectories (which were highly tangled) were less well fit by a linear dynamical model (R^2 = 0.51 and 0.37 for monkey D and C) than were the empirical neural trajectories (R^2 = 0.79 and 0.73). Despite this agreement, we avoided using the above R^2 as our primary measure, because there exist trajectories that could be readily produced by a dynamical system with smooth dynamics but are poorly described by a linear model – e.g., the trajectory in Figure 7A (*right subpanel*). We also found that the quality of a linear dynamical fit was somewhat sensitive to both the span of time and the number of dimensions considered. In contrast, tangling gave consistent results regardless of such choices.

Computational Motivation for the Tangling Measure

Here we show that, given limits on how rapidly a flow-field can change, when two trajectories (or two portions of the same trajectory) come close and then diverge, a potential instability is inevitable. We define a potential instability as a direction along which an error will grow with time in the local vicinity. The argument below is a simple proof by contradiction. Avoiding a potential instability requires that, for all directions, local errors shrink with time. For a linearized system, this implies that all eigenvalues are less than zero. Yet if two trajectories diverge, there must be at least one positive eigenvalue.

Assume two time-evolving trajectories, $\boldsymbol{x}_1(t)$, and $\boldsymbol{x}_2(t')$. These could be two portions of a larger trajectory or could correspond to two different conditions. We consider the moment where they become closest: i.e., when $\|\boldsymbol{x}_1(t) - \boldsymbol{x}_2(t')\|$ is smallest. Without loss of generality, we assume this happens at $t = 0$ and $t' = 0$. We also consider the state, $\bar{\boldsymbol{x}}$ halfway between $\boldsymbol{x}_1(0)$ and $\boldsymbol{x}_2(0)$. Without loss of generality, we define $\bar{\boldsymbol{x}}$ as the origin. Thus $\boldsymbol{x}_1(0) = -\boldsymbol{x}_2(0)$. As in Figure S1, we assume that tangling between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is high because $\|\dot{\boldsymbol{x}}_1(0) - \dot{\boldsymbol{x}}_2(0)\|$ is large while $\|\boldsymbol{x}_1(0) - \boldsymbol{x}_2(0)\|$ is small. We can therefore use the Taylor series to approximate the flow-field at state $\boldsymbol{x}$ in the vicinity of $\bar{\boldsymbol{x}}$. We ignore higher-order terms:

$$\dot{\boldsymbol{x}} = \boldsymbol{a} + B\boldsymbol{x}$$

where the matrix B is the Jacobian evaluated at $\boldsymbol{x} = 0$.

Because both $\boldsymbol{x}_1(0)$ and $\boldsymbol{x}_2(0)$ are near $\bar{\boldsymbol{x}}$, we have:

$$\dot{\boldsymbol{x}}_1(0) = \boldsymbol{a} + B\boldsymbol{x}_1(0)$$

and

$$\dot{\boldsymbol{x}}_2(0) = \boldsymbol{a} + B\boldsymbol{x}_2(0) = \boldsymbol{a} - B\boldsymbol{x}_1(0).$$

We now consider some perturbation of the $\boldsymbol{x}_1$ trajectory, such that $\boldsymbol{x}_1'(0) = \boldsymbol{x}_1(0) + \boldsymbol{\varepsilon}$. Stability requires, $\forall \boldsymbol{\varepsilon}$:

$$\begin{aligned}
&\|\boldsymbol{x}_1'(\Delta t) - \boldsymbol{x}_1(\Delta t)\|^2 < \|\boldsymbol{x}_1'(0) - \boldsymbol{x}_1(0)\|^2 \\
&\Rightarrow \|\left(\boldsymbol{x}_1'(0) + \Delta t\left(\boldsymbol{a} + B\boldsymbol{x}_1'(0)\right)\right) - \left(\boldsymbol{x}_1(0) + \Delta t(\boldsymbol{a} + B\boldsymbol{x}_1(0))\right)\|^2 < \|\boldsymbol{x}_1(0) + \boldsymbol{\varepsilon} - \boldsymbol{x}_1(0)\|^2 \\
&\Rightarrow \|\boldsymbol{\varepsilon} + \Delta t B\boldsymbol{\varepsilon}\|^2 < \|\boldsymbol{\varepsilon}\|^2 \\
&\Rightarrow \|\boldsymbol{\varepsilon}\|^2 + 2\Delta t\boldsymbol{\varepsilon}^{\mathrm{T}}B\boldsymbol{\varepsilon} + \Delta t^2\boldsymbol{\varepsilon}^{\mathrm{T}}B^{\mathrm{T}}B\boldsymbol{\varepsilon} < \|\boldsymbol{\varepsilon}\|^2 \\
&\Rightarrow \|\boldsymbol{\varepsilon}\|^2 + 2\Delta t\boldsymbol{\varepsilon}^{\mathrm{T}}B\boldsymbol{\varepsilon} < \|\boldsymbol{\varepsilon}\|^2, \text{ as } \Delta t^2 \text{ is very small.} \\
&\Rightarrow \boldsymbol{\varepsilon}^{\mathrm{T}}B\boldsymbol{\varepsilon} < 0
\end{aligned}$$

Because this must be true for all $\boldsymbol{\varepsilon}$, this is equivalent to stating that all eigenvalues of B must be negative. However, because $\boldsymbol{x}_1(t)$, and $\boldsymbol{x}_2(t)$ are closest at $t = 0$, we have:

$$\begin{aligned}
&\|\boldsymbol{x}_1(\Delta t) - \boldsymbol{x}_2(\Delta t)\|^2 > \|\boldsymbol{x}_1(0) - \boldsymbol{x}_2(0)\|^2 \\
&\Rightarrow \|(\boldsymbol{x}_1(0) + \Delta t(\boldsymbol{a} + B\boldsymbol{x}_1(0))) - (\boldsymbol{x}_2(0) + \Delta t(\boldsymbol{a} + B\boldsymbol{x}_2(0)))\|^2 > \|\boldsymbol{x}_1(0) - \boldsymbol{x}_2(0)\|^2 \\
&\Rightarrow \|2\boldsymbol{x}_1(0) + 2\Delta t B\boldsymbol{x}_1(0)\|^2 > \|2\boldsymbol{x}_1(0)\|^2 \\
&\Rightarrow \|\boldsymbol{x}_1(0)\|^2 + 2\Delta t\boldsymbol{x}_1(0)^{\mathrm{T}}B\boldsymbol{x}_1(0) + \Delta t^2\boldsymbol{x}_1(0)^{\mathrm{T}}B^{\mathrm{T}}\boldsymbol{x}_1(0) > \|\boldsymbol{x}_1(0)\|^2 \\
&\Rightarrow \|\boldsymbol{x}_1(0)\|^2 + 2\Delta t\boldsymbol{x}_1(0)^{\mathrm{T}}B\boldsymbol{x}_1(0) > \|\boldsymbol{x}_1(0)\|^2, \text{ as } \Delta t^2 \text{ is very small.} \\
&\Rightarrow \boldsymbol{x}_1(0)^{\mathrm{T}}B\boldsymbol{x}_1(0) > 0
\end{aligned}$$

This is in contradiction to the claim above that $\boldsymbol{\varepsilon}^{\mathrm{T}}B\boldsymbol{\varepsilon} < 0$ for $\forall\ \boldsymbol{\varepsilon}$. Equivalently, it implies that at least one eigenvalue of B must be positive, in contrast to the claim above that all eigenvalues must be negative.

Thus, local stability is inconsistent with the fact that trajectories are close but diverging. The above argument does not strictly depend on $\|\dot{\boldsymbol{x}}_1(0) - \dot{\boldsymbol{x}}_2(0)\|$ being large. However, a larger $\|\dot{\boldsymbol{x}}_1(0) - \dot{\boldsymbol{x}}_2(0)\|$ implies larger positive eigenvalue(s) of B. All other things being equal, this will result in a larger potential instability due to greater local divergence.

Standard Recurrent Neural Networks

We used two very different approaches to train recurrent neural networks (RNNs). In the first approach, we trained RNNs to produce a target output (Figure 5) as is conventionally done. We used a network with dynamics:

$$\boldsymbol{x}(t+1, c) = f(A\boldsymbol{x}(t, c) + B\boldsymbol{u}(c) + \boldsymbol{w}(t, c))$$

where $\boldsymbol{x}$ is the network state (the 'firing rate' of every unit) for time t and condition c. The function $f := \tanh$ is an element-wise transfer function linking a unit's input to its firing rate, $A\boldsymbol{x}$ captures the influence of network activity on itself via the connection weights in A, $B\boldsymbol{u}$ captures external inputs, and the random vector $\boldsymbol{w} \sim N(0, \sigma_w I)$ adds modest noise. Network output is then a linear readout of its firing rates:

$$\boldsymbol{y}(t, c) = C\boldsymbol{x}(t, c)$$

The parameters A, B, C, and $\boldsymbol{x}(0, c)$ were optimized to minimize the difference between the network output, $\boldsymbol{y}$ and a target, $\boldsymbol{y}_{\text{targ}}$. That target output was the pattern of activity, across all muscles, during the middle five cycles of a seven-cycle movement. We used

two conditions with different target outputs: $\boldsymbol{y}_{\text{targ}}(:,1)$ and $\boldsymbol{y}_{\text{targ}}(:,2)$ contained muscle activity during forward and backward cycling respectively. The input provided the network with the condition identity: $\boldsymbol{u}(1)=[1;0]$ and $\boldsymbol{u}(2)=[0;1]$.

The loss function optimized during training contained both error and regularization terms:

$$L=\sum_{t,c}\left[\frac{1}{2}\|\boldsymbol{y}_{\text{targ}}(t,c)-\boldsymbol{y}(t,c)\|_2^2\right]+\frac{\lambda_A}{2}\|A\|_F^2+\frac{\lambda_C}{2}\|C\|_F^2+\sum_{t,c}\left[\frac{\lambda_x}{2}\|\boldsymbol{x}(t,c)\|_2^2\right]$$

where the first term is the error between the network output and the target, the second and third terms penalize large recurrent and output weights respectively, and the last term penalizes large firing rates. By varying the hyper-parameters λ_A, λ_C, λ_x, σ_w, and the initial weight values, we simulated a family of networks that found different solutions for producing the same output. This allowed us to ask whether low network-trajectory tangling was a common feature of those solutions.

We trained 1000 such networks. Hyper-parameters were drawn randomly from log uniform distributions, $\lambda_A\in[10^{-4},10^{-1}]$, $\lambda_C\in[10^{-6},10^{1}]$, $\lambda_x\in[10^{-4},10^{1}]$, and $\sigma_w\in[10^{-4},10^{1}]$. Each RNN included $n=100$ units. Each matrix of the RNN was initialized to a random orthonormal matrix. RNNs were trained using TensorFlow's Adam optimizer. We discarded RNNs that were not successful ($R^2<0.5$ between target and actual outputs). Because of the broad range of hyper-parameters, only a subset of networks (463) were successful.

As a technical point, we were concerned that, despite regularization, networks might find overly specific solutions. Each cycle of the empirical muscle activity had different small idiosyncrasies, and optimization might promote overfitting of these small differences. We therefore added 'new' conditions to $\boldsymbol{y}_{\text{targ}}(t,c)$. Each new condition involved a target output that was almost identical to that for one of the original two conditions, but was modified such that the small idiosyncrasies occurred on different cycles. This ensured that networks produced a consistent output very close to the empirical muscle activity, but did not attempt to perfectly match small cycle-specific idiosyncrasies. The inclusion of noise via $\boldsymbol{w}$ also encouraged optimization to find robust, rather than overfit, solutions. Noise magnitude, σ_w, was a hyper-parameter that was varied across networks, to encourage varied solutions. However, σ_w was always set to zero when measuring network tangling.

Trajectory-Constrained Neural Networks

To examine how tangling relates to noise-robustness (Figure 7B) we trained RNNs to follow a set of target internal trajectories. This involved the unconventional approach of employing both a target output, $\boldsymbol{y}_{\text{targ}}$, and a target internal network trajectory, $\boldsymbol{s}_{\text{targ}}$. Networks consisted of 100 units. Network dynamics were governed by

$$\boldsymbol{v}(t+1)=\boldsymbol{v}(t)+\Delta t/\tau(-\boldsymbol{v}(t)+A\,f(\boldsymbol{v}(t))+\boldsymbol{w}(t))$$
$$\boldsymbol{y}(t)=Cf(\boldsymbol{v}(t))$$

where $f:=\tanh$, and $\boldsymbol{w}\sim N(0,\sigma_w I)$ adds noise. $\boldsymbol{v}$ can be thought of as the membrane voltage and $f(\boldsymbol{v}(t))$ as the firing rate. $Af(\boldsymbol{v}(t))$ is then the network input to each unit: the firing rates weighted by the connection strengths. $Cf(\boldsymbol{v}(t))$ is a linear readout of firing rates.

During training, A was adjusted using recursive least-squares (Sussillo and Abbott, 2009) so that $Af(\boldsymbol{v}(t))\approx\boldsymbol{s}_{\text{targ}}$. Training thus insured that the synaptic inputs to each unit closely followed the pre-determined trajectory defined by $\boldsymbol{s}_{\text{targ}}$. Firing rates therefore also followed a pre-determined trajectory. C was adjusted so that $\boldsymbol{y}\approx\boldsymbol{y}_{\text{targ}}$. Training was deemed successful if the R^2 between $\boldsymbol{y}$ and $\boldsymbol{y}_{\text{targ}}$ was >0.9. Noise tolerance was assessed as the largest value of σ_w for which the network could be trained to accurately produce the target output for five consecutive cycles ($R^2>0.9$ between $\boldsymbol{y}$ and $\boldsymbol{y}_{\text{targ}}$, averaged across 100 iterations) despite the constraint of following the target internal trajectory, $\boldsymbol{s}_{\text{targ}}$.

We set $\boldsymbol{y}_{\text{targ}}=[\cos t;\ \sin 2t]$. To construct $\boldsymbol{s}_{\text{targ}}$, we began with an idealized low-dimensional target, $\boldsymbol{s}(t)'_{\text{targ}}=[\cos t;\ \sin 2t;\ \beta\sin t]$. To give each unit a target, we set $\boldsymbol{s}_{\text{targ}}=G\boldsymbol{s}'_{\text{targ}}$ where G is a random matrix of size 100×3 with entries drawn independently from a uniform distribution from -1 to 1. Noise tolerance was tested for a range of values of β. That range produced target trajectories that varied greatly in their tangling, allowing us to examine how tangling related to noise tolerance. Noise tolerance was the largest magnitude of state noise for which the network still produced the desired output. For each target trajectory, and each of the 20 random initializations of A, C, and G, we doubled σ_w starting at 0.005 until we found the noise tolerance. We then computed the average (and SEM) noise tolerance across the 20 parameter initializations.

Predicting Neural Population Activity

The optimization described by Equation 2 was performed using the Theano Python module. Optimization was initialized either with $\widehat{X}_{init}=Z$, or with $\widehat{X}_{init}=Z+noise$ where the noise was smooth with time but independent for each dimension. Both $\widehat{X}$ and Z were $10\times T$; they contained the projection onto the top ten PCs. T is the total number of time points across the conditions being considered. Specifically, we predicted neural activity for three middle cycles of forward cycling and three middle cycles of backward cycling (both taken from seven-cycle movements). Because dimensionality is equal for $\widehat{X}$ and Z, the ability to decode Z from $\widehat{X}$ will suffer as optimization modifies $\widehat{X}$. However, because some dimensions of Z contain more variance than others, $\widehat{X}$ can gain considerable new structure while compromising the decode only modestly. This tradeoff can be determined by the choice of λ. However, for scientific reasons, we employed a modified approach to better control that tradeoff. We wished to ensure that the predictions made by different cost functions all encoded muscle activity equally well. This aids interpretation when comparing the results of the

optimization in Figures 7C and 7D with optimizations using different cost functions in Figure S7. By matching encoding accuracy, any differences in similarity must be due to other structure that differs due to the cost function being optimized. Thus, instead of minimizing the first term of Equation 2 (which attempts to create a perfect decode) we minimized the squared difference between the decode R^2 and 0.95. We only considered optimizations that achieved this with a tolerance of 0.01. This approach insures that muscle encoding is equally good for the predicted populations responses yielded by different cost functions. Optimizations employed gradient descent using an inexact line search for the Wolfe conditions $c_1 = 0.05$ and $c_2 = 0.1$. As a technical point, the derivative used to compute $Q(t_{end})$ was based on the assumption that the three-cycle pattern would repeat.

Similarity between Empirical and Predicted Data

We assessed similarity using a modified version of canonical correlation (Cunningham and Ghahramani, 2015). This method finds a pair of orthogonal transformations, one for each dataset, that maximizes the correlation between the transformed datasets. Specifically, for mean-centered datasets $X_a \in \mathbb{R}^{KxT}$ and $X_b \in \mathbb{R}^{KxT}$, similarity is:

$$S(X_a, X_b) = \underset{M_a, M_b}{\text{argmax}} \frac{tr\left(M_a^\top X_a X_b^\top M_b\right)}{\sqrt{tr\left(M_a^\top X_a X_a^\top M_a\right) tr\left(M_b^\top X_b X_b^\top M_b\right)}}.$$

Subject to the constraint that M_a and M_b are orthonormal matrices. Similarity will thus be unity if two datasets are the same but for an orthonormal transformation. Note also that an overall shift of one dataset relative to the other does not impact similarity because the data are mean-centered before computing similarity. Due to the normalization in the denominator of the above cost function, similarity is also not impacted by an isotropic scaling of one dataset relative to the other.

Predictions via Alternate Cost Functions

We performed a variety of optimizations corresponding to several alternate cost functions (Figure S7). Each cost function embodied a hypothesis regarding the relationship between neural activity and muscle activity.

All cost functions were of the form:

$$\hat{X} = \underset{X}{\text{argmin}} \sum_{k=1}^{K} \lambda_k f_k(X, Z)$$

where f_k is some function of the input data and λ_k are scaling coefficients used to ensure that one term of the cost function did not dominate at the expense of the others. The arguments of $f_k()$ are the optimization variable, X and the empirical muscle activity, Z. All cost functions examined in Figure S7 are described below in terms of different definitions of $f_k()$.

Muscle encoding and low tangling (same as Equation 2):

$$f_1(X, Z) = f_{\text{decode}}(X, Z) = \| Z - ZX^\dagger X \|_F^2$$
$$f_2(X) = f_{\text{tangling}}(X) = \sum_t Q_X(t)$$

Nonlinear mapping with L-2 minimization:

$$f_1\left(X, \overline{Z}\right) = f_{\text{decode-nonlin}}\left(X, \overline{Z}\right) = \|\overline{Z} - \hat{\overline{Z}}\|_F^2$$

$\overline{Z}$ contains individual muscle activity. Here we consider the activity of all muscles individually (rather than the top ten PCs as above) because this matters in the non-linear case. The hypothesis being considered is that motor cortex may use a simplified set of muscle 'synergies' that becomes, via a set of non-linear transformations, the activity of each muscle. $\hat{\overline{Z}} = \boldsymbol{\alpha} + \tanh(BX + \boldsymbol{\gamma})$ with the parameters $\boldsymbol{\alpha}$, B, and $\boldsymbol{\gamma}$ optimized to minimize $f_{\text{decode-nonlin}}(X, \overline{Z})$.

$$f_2(X) = f_{\text{norm}}(X) = \|X\|_F^2$$

where F denotes the Frobenius norm.

Nonlinear mapping with tangling minimization:

$$f_1\left(X, \overline{Z}\right) = f_{\text{decode-nonlin}}\left(X, \overline{Z}\right)$$

$$f_2(X) = f_{\text{tangling}}(X)$$

where $f_{\text{decode-nonlin}}$ and f_{tangling} are as described above.

Low curvature:

$$f_1(X, Z) = f_{\text{decode}}(X, Z)$$

$$f_2(X) = f_{curvature}(X) = \sum_t \frac{\| \dot{\boldsymbol{x}}_t^{\text{norm}} - \dot{\boldsymbol{x}}_{t-1}^{\text{norm}} \|}{s_t}$$

where,

$$\dot{\boldsymbol{x}}_t^{\text{norm}} = \frac{\dot{\boldsymbol{x}}_t}{\| \dot{\boldsymbol{x}}_t \|}$$

and s_t is the normalized 'speed' of the neural trajectory,

$$s_t = \frac{\| \dot{\boldsymbol{x}}_t \|}{\sum_{t'} \| \dot{\boldsymbol{x}}_{t'} \|}$$

As a technical point, we wished to ensure that the predictions made by different cost functions all encoded muscle activity equally well. By matching the accuracy of muscle encoding, any differences in similarity must be due to other structure introduced during optimization. We therefore modified $f_{\text{decode}}(X, Z)$ and $f_{\text{decode}-\text{nonlin}}(X, \overline{Z})$ so that they were minimized when decode accuracy had an R^2 of 0.95, rather than 1.0. We only considered optimizations that achieved this with a tolerance of 0.01.