Dynamic Encoding of Movement Direction in Motor Cortical Neurons

Jörn Rickert,1,2 Alexa Riehle,4 Ad Aertsen,1,2 Stefan Rotter,2,3 and Martin P. Nawrot2,5
1Neurobiology and Biophysics, Faculty of Biology, 2Bernstein Center for Computational Neuroscience Freiburg, and 3Computational Neuroscience, Faculty of Biology, Albert Ludwigs University, 79104 Freiburg, Germany, 4Mediterranean Institute of Cognitive Neuroscience–Centre National de la Recherche Scientifique and University of Aix-Marseille, 13402 Marseille, France, and 5Neuroinformatics and Theoretical Neuroscience, Institute of Biology, Freie Universität Berlin, and Bernstein Center for Computational Neuroscience Berlin, 14195 Berlin, Germany

When we perform a skilled movement such as reaching for an object, we can make use of prior information, for example about the location of the object in space. This helps us to prepare the movement, and we gain improved accuracy and speed during movement execution. Here, we investigate how prior information affects the motor cortical representation of movements during preparation and execution. We trained two monkeys in a delayed reaching task and provided a varying degree of prior information about the final target location. We decoded movement direction from multiple single-unit activity recorded from M1 (primary motor cortex) in one monkey and from PMd (dorsal premotor cortex) in a second monkey. Our results demonstrate that motor cortical cells in both areas exhibit individual encoding characteristics that change dynamically in time and dependent on prior information. On the population level, the information about movement direction is at any point in time accurately represented in a neuronal ensemble of time-varying composition. We conclude that movement representation in the motor cortex is not a static one, but one in which neurons dynamically allocate their computational resources to meet the demands defined by the movement task and the context of the movement. Consequently, we find that the decoding accuracy decreases if the precise task time, or the previous information that was available to the monkey, were disregarded in the decoding process. An optimal strategy for the readout of movement parameters from motor cortex should therefore take into account time and contextual parameters.

Introduction

Neurons in motor cortex are well known to encode physical parameters of limb movements such as direction, force, extent, load, or posture (for review, see Hepp-Reymond, 1988; Johnson et al., 2001). In a series of experimental studies, it has been shown that encoding properties are sufficiently stable over time, such that arm movement trajectories could be accurately decoded from neural populations in single trials (Wessberg et al., 2000; Serruya et al., 2002; Taylor et al., 2002; Mehring et al., 2003). This opened the possibility to decode neural spike trains in brain–machine interfacing (BMI), as a means to control prosthetic devices (Carmena et al., 2003; Musallam et al., 2004; Hochberg et al., 2006; Santhanam et al., 2006; Moritz et al., 2008; Velliste et al., 2008). Indeed, real-time applications have, in principle, proven to work under experimentally controlled conditions. However, several recent studies provided evidence that neural encoding properties need not be stable under all circumstances. Instead, they can change dynamically with time, both within a trial and across repeated trials (Chestek et al., 2007; Churchland and Shenoy, 2007; Rokni et al., 2007). Such changes may be attributable to changes in contextual parameters. In parietal cortex, the influence of attention (Oristaglio et al., 2006), reward (Musallam et al., 2004), or movement cue (Gail and Andersen, 2006) on the neural encoding of movement goals is well established (for review, see Gottlieb, 2007). In motor cortex, there is evidence for similar influences on the encoding of movements. For instance, reward and reward expectation can modulate motor cortical activity related to eye movements (Roesch and Olson, 2004). Serial execution order of different movement components influences their neural representation (Carpenter et al., 1999; Hepp-Reymond et al., 1999; Lu and Ashe, 2005), as well as the presence or absence of specific movement cues (Hepp-Reymond et al., 1999). Moreover, motor cortical activity can be similar during actual execution and during the preparation of a movement (for review, see Riehle, 2005).

Here, we studied in detail the time-dependent encoding of movement direction in the monkey primary motor cortex (M1) and dorsal premotor cortex (PMd) during movement preparation and execution in a delayed hand-reaching task with a varying context. In different experimental conditions, two monkeys were confronted with different degrees of previous target information, presented at the beginning of the initial hold period. Previous
studies showed that prior information modulates motor cortical activity during movement preparation (Georgopoulos et al., 1989; Riehle and Requin, 1989; Bastian et al., 2003; Roux et al., 2003; Cisek and Kalaska, 2005; Churchland et al., 2006) and during execution (Riehle and Requin, 1989; Hepp-Reymond et al., 1999). Our analyses, using single-trial Bayesian inference, show that the encoding of movement direction in M1 (monkey 1) and in PMd (monkey 2) is a dynamic process dependent on time and prior information. Nevertheless, on the population level, 50–100 neurons allowed an essentially error-free decoding of the visually presented cue and of the executed movement direction at any point in time during movement preparation and execution.

Materials and Methods

Experimental paradigm. Two monkeys (*Macaca mulatta*), one male (monkey 1) and one female (monkey 2), were trained to perform a delayed arm-reaching task. The animals were cared for according to European and French government regulations. They were seated in front of a vertical panel carrying seven touch-sensitive buttons with light-emitting diodes (LEDs)—one in the center and six on a circle of 10 cm radius around it (see Fig. 1). When the center LED was lit, the monkey had to place its right hand on it to initiate the trial (trail start (TS)). After a delay of 300 ms, one, two, or three adjacent target LEDs lit up in green (preparatory signal (PS)), providing the monkey with more or less information about the direction of the upcoming movement (see Fig. 1A): (1) in the one-target condition, one of the six target LEDs lit up; (2) in the two-target condition, one of three possible pairs of adjacent target LEDs lit up (targets 1 + 2, 3 + 4, or 5 + 6); (3) in the three-target condition, one of two possible triplets of adjacent target LEDs lit up (targets 6 + 1 + 2 or 3 + 4 + 5). After a fixed preparatory period of one second, one of the green LEDs turned red (response signal (RS)), signaling the monkey to release the center LED and to reach out and touch the red target signal. Thus, each condition comprised six types of trials (i.e., six possible combinations of the preparatory and the response signal). Correct and fast performance was rewarded by fruit juice. The three behavioral conditions of prior information (i.e., one, two, or three targets as preparatory signal) were presented as separate blocks of ~100 trials each, within which each trial type was presented at random with equal probability (median of 19 trials per final target direction with a SD of 6.1 in monkey 1 and 13 trials per final target direction with a SD of 4.2 in monkey 2). On each recording day, the order of blocks was chosen randomly. Mean reaction times were 161, 195, and 204 ms for monkey 1, and 234, 236, and 237 ms for monkey 2, for the one-, two-, and three-target conditions, respectively. Mean movement duration was 199, 202, and 209 ms for monkey 1, and 181, 177, and 184 ms for monkey 2 (for additional details on the relationship of reaction time and prior information, see Bastian et al., 2003). To simplify terminology, the 1 s period between PS and RS was termed “preparatory period” and a 300 ms period from RS to ~100 ms before movement end was termed “execution period.”

Electrophysiological recordings. After the monkey had successfully learned the task (>85% correct performance), a cylindrical recording chamber (inner diameter, 15 mm) was implanted over the contralateral M1 and the PMd between central sulcus (CS), the posterior bank of the arcuate sulcus, and the precentral dimple, under aseptic conditions and general halothane anesthesia (<2.5% in air). In monkey 1, the recording chamber was positioned close to the CS (see Fig. 1B) and electrodes were positioned in M1; in monkey 2, the chamber position was more frontal (see Fig. 1C) mainly over PMd. A stainless-steel T-bar was cemented to the skull to fixate the animal’s head during recording sessions. A multielectrode microdrive (Reitböck system; Thomas Recording) was used to transdurally insert up to seven quartz insulated platinum–tungsten electrodes (outer diameter, 80 μm; impedance, 2–5 MΩ at 1 kHz). Electrodes were spaced ~300 μm apart from each other. The use of a window discriminator yielded the spike times of up to two neurons per electrode. The neuronal signals along with behavioral events (trial start and end, target information, reaction and movement times, reward, errors) were stored for off-line analysis at a time resolution of 1 ms. Here, we used only correct trials for analysis.

Estimation of neuronal firing rates. In most cases (see Figs. 2–7A), spike trains were aligned to the occurrence of PS (for an example, see Fig. 2A). Neuronal firing rate profiles were estimated from single-trial spike trains (Navròt et al., 1999) by convolution with a fixed Gaussian kernel of a standard width of 65 ms and area normalized to 1 (see example in Fig. 2B). The exact width of the kernel was not critical to our results. Kernels of widths in the range of 20–100 ms lead to the same general conclusions.

Measures for directional tuning. We considered movements into six different movement directions d that were executed after RS. They were specified by the six uniformly arranged unit vectors \(\vec{e}_d\). For a sextet of neuronal firing rate profiles \(f_r(t)\), one for each direction \(d\), a fundamental quantity is given by the weighted vectorial sum as follows:

\[
\bar{a}(t) = \sum_d f_r(t) \vec{e}_d. \tag{1}
\]

At each point in time, the resulting vector \(\bar{a}(t)\) represents the instantaneous directional tuning. Its length \(|\bar{a}(t)|\) describes the tuning amplitude, or directional modulation. Its direction is specified by the unit vector \(\hat{a}(t)/|\bar{a}(t)|\) or, alternatively, by the angle \(\phi(t)\) with respect to the upward direction. Let \(\langle \cdot \rangle\) denote trial averages. The trial-averaged tuning vector \(\bar{A}(t)\) is the vector mean computed from the direction-specific trial-averaged firing rates \(R_d(t) = \langle f_r(t)\rangle\) as follows:

\[
\bar{A}(t) = \langle r(t) \rangle = \sum_d R_d(t) \vec{e}_d. \tag{2}
\]

Let \(A(t)\) denote the length of the trial-averaged tuning vector \(\bar{A}(t)\). The variance of tuning can be predicted from the variances of the rates \(f_r(t)\), provided that the latter are uncorrelated across directions. Two biologically relevant aspects of tuning variance are used to assess the reliability of directional tuning:

1. Let \(\sigma_z^2(t) = \langle [\bar{a}(t) - \bar{A}(t)]^2 \rangle\) (3)

\[
\sigma_z^2(t) = \langle [\bar{a}(t) - \bar{A}(t)]^2 \rangle. \tag{3}
\]

\(\sigma_z^2(t)\) is used as a measure for the reliability of the directional tuning.

2. Circular statistics (Batschelet, 1981) allowed us to evaluate the preferred direction without regard of the tuning amplitude. Let \(\bar{A}_d(t) = \langle \bar{a}(t)/|\bar{a}(t)|\rangle\) (5)

\[
\bar{A}_d(t) = \langle \hat{a}(t)/|\bar{a}(t)|\rangle. \tag{5}
\]

denote the mean preferred direction; \(A_d(t)\), its length; and \(\Phi(t)\), its angle with respect to the upward direction. The fact that \(\bar{A}_d(t) = \langle \cos(\phi(t) - \Phi(t)) \rangle\)

\[
\bar{A}_d(t) = \langle \cos(\phi(t) - \Phi(t)) \rangle \tag{6}
\]

justifies the use of \(\phi(t) - \Phi(t)\) as a measure for the angular variance. Its square root is called angular deviation of the tuning vector from its mean). The \(z\) score relating the length of the tuning vector to its scalar variance

\[
z_a(t) = A(t)/\sigma_z(t) \tag{4}
\]

• 13871

J. Neurosci., November 4, 2009 • 29(44):13870 –13882 • 13871

Rickert et al. • Dynamic Movement Encoding in MC
Bayesian inference of movement direction. We wanted to know how accurately the movement direction executed after RS is represented in single-trial firing rates throughout the trial. To measure this, we used Bayesian inference from nonparametric density estimations. This method of Bayesian decoding assumes the perspective of an “ideal observer” (for review, see Pouget et al., 2000; Ma et al., 2006). It yields an estimate of the information content of neuronal activity without specifying the encoding strategy that is used by the brain. Our approach makes two constraining assumptions: (1) Firing rate is a sufficient descriptor of neuronal activity (i.e., the precise timing of action potentials is not considered to represent information), and (2) firing rates are assumed to be uncorrelated across neurons. The latter assumption enabled us to analyze pseudopopulations of neurons, most of them recorded in different sessions, but during trials in the same experimental condition.

In detail, this method constitutes the following steps: Trial-averaged tuning curves were replaced by probabilistic tuning profiles in terms of the following conditional probability:

\[ P(r | d) : \text{directional tuning of rate.} \]

Such probabilities are conveniently represented by appropriately normalized probability density functions. Estimates of these were obtained by kernel convolution with reflection to account for the nonnegativity of rates (Silverman, 1986). To assess the suitability of such a tuning model for the decoding of neuronal activity, we again used conditional probabilities, but now with the roles of \( r \) and \( d \) reversed:

\[ P(d | r) : \text{inference of direction from rate} \]

measures the “plausibility” of a movement in direction \( d \), given a specific observation of neuronal firing rate \( r \). The Bayes rule was used to compute plausibilities from tuning profiles (Johannesma, 1981; Gelman et al., 1995; Pouget et al., 2000) as follows:

\[
    P(d | r) = \frac{P(d) P(r | d)}{P(r)} = P(r | d), \quad (8)
\]

For any particular single-trial rate profile \( r_d(t) \) recorded before and during a movement to direction \( d \), the corresponding plausibility profile can be read out. This leads to six distinct probability profiles as follows:

\[ P(d' | r_d(t)) \]

measuring the plausibility of direction \( d' \) if the rate profile \( r_d(t) \) was observed. The expectation over 20 randomly chosen trials (in very few cases with <20 trials, some were used twice) for each movement direction with \( d' = d \) gives the average probability

\[
    P_d(t) = \langle P(d | r_d(t)) \rangle, \quad (9)
\]

for the correct decoding at time \( t \) of the direction \( d \) of the actual movement that was performed. Note that the direction of the actual movement was in all cases identical with the direction indicated by the target cue presented at RS since we considered only correct trials. The training trials used to estimate these probabilities from experimental data did not include the respective test trial for which the plausibility profiles were compiled (cross-validation, leave-one-out) (Efron and Tibshirani, 1993). This provided us with an efficient means to assess the expected single-trial readout performance for a particular neuron; we refer to this measure as the decoding probability.

To compute whether the decoding probability of a neuron lies significantly above chance level, we compared the decoding probability to a chance distribution of decoding probabilities computed from a flat tuning curve. The chance distribution was obtained numerically by computing 100 times the decoding probability from single-trial rates of the same neuron but the rates were randomly sorted into six groups. To limit computing time, decoding probabilities at any time during the preparatory period were compared with one chance distribution computed from the mean firing rates of the neurons during the preparatory period. Likewise, decoding probabilities any time during the execution period were compared with one chance distribution from the mean rates during the execution period.

Next, we considered a neuronal population comprising \( n \) neurons. The joint decoding of its activity given by the rates \( r_1, r_2, \ldots, r_n \) should be based on the joint tuning curves as follows:

\[ P(r_1, r_2, \ldots, r_n | d) : \text{joint directional tuning of rates.} \]

The assumption of independent encoding allows us to write the following:

\[ P(r_1, r_2, \ldots, r_n | d) = P(r_1 | d) \cdot P(r_2 | d) \cdot P(r_3 | d) \cdot \ldots \cdot P(r_n | d). \quad (10) \]

Based on these joint probabilities, we can deal with neuronal populations in very much the same way as we did with single neurons.

Furthermore, we computed decoding probabilities based on different combinations of behavioral conditions (one, two, and three targets) used for training and test trials: (1) “context known”: same condition used for training and test trials; (2) “context unknown”: all three available conditions used for training trials; (3) “generalization”: training trials taken from a single condition different from the test trials condition.

In formal terms, using the Bayes probability for multiple variables, Equation 8 in our computation has to be replaced by the following:

\[
    P(d | r, \text{cond}_a) = \frac{P(d) \cdot P(r | d, \text{cond}_a)}{P(r | \text{cond}_a)}. \quad (11)
\]

In our case, all conditions could be considered equally likely, such that the equation simplifies to the following:

\[
    P(d | r, \text{cond}_a) = \frac{P(d) \cdot P(r | d, \text{cond}_a) \cdot P(r | \text{cond}_a)}{P(r | \text{cond}_a)}. \quad (12)
\]

This reflects the context known case and is exactly what was used to compute the decoding probabilities in Figures 3–7A. The other two cases are reflected by the following two equations:

\[ P(d | r, \text{cond}_a) = \frac{P(d) \cdot P(r | d, \text{cond}_a) \cdot P(r | \text{cond}_a)}{P(r | \text{cond}_a)}. \quad (12)\]
was used as the estimated task time.

The point in time for which the integrated squared error was minimal

where training data from another condition are used to approximate the probability

of a direction when the context is unknown, for the context unknown case, and

where training data from another condition are used to approximate the probability

of a direction with a different context, for the generalization case.

In the procedure described above, throughout the trial we inferred the movement direction that was instructed with RS and executed shortly after (only correct trials). For the two- and three-target conditions, target information was incomplete during the 1 s preparatory period. In the three-target (two-target) condition, the monkey could not know which of the three (two) possible targets would light up with the RS. Thus, before RS we expect that the maximally possible decoding probability for the final target is $P_c = 1/3$ ($P_c = 1/2$), given that there was no bias for any specific target in our experimental design.

Inference of the preparatory stimulus. Alternatively, we computed the neuronal representation of the stimulus (rather than that of the final movement direction). Here, the probability of observing stimulus $s'$ given the firing rates $r$, was calculated using the same formal approach that led to Equations 9 and 12. In the two- and three-target conditions, we have three or two preparatory stimuli (PSs), and thus the decoding is reduced to a three- or two-class problem, respectively. The maximally possible decoding probability for the stimulus now is $P_r = 1$ for all stimuli, regardless of the condition. These results were qualitatively similar to those for the decoding of the final movement. They are included in the supplemental material (available at www.jneurosci.org).

Inference of task time. We estimated the task time (in the range from 300 ms before PS to 300 ms after RS) from single-trial firing rates computed as described above. From a test data vector of single-trial firing rates (computed as described above), the maximally possible decoding probability for the stimulus now is $P_r = 1$ for all stimuli, regardless of the condition. These results were qualitatively similar to those for the decoding of the final movement. They are included in the supplemental material (available at www.jneurosci.org).

The time-dependent directional tuning of the mean firing rate of a neuron was visualized with the tuning vector (i.e., the vectorial sum of the trial-averaged firing rates from each movement direction executed after RS) (see Materials and Methods). During the time course of the trial, the resulting vector plot expresses the temporal evolution of both, the amplitude of tuning (vector length), and the preferred direction (vector direction) (for examples, see Fig. 3A, top; B, C, top panels). Before PS, the monkey had no information about the movement target, and thus, neuronal activity could not have been tuned to target direction. This resulted in very small residual amplitudes and random directions of the tuning vector, reflecting the estimation bias attributable to limited sample size (Fig. 3A–C, top panels). After PS, the vector amplitude increased and the vector direction stabilized as the tuning of the neuron evolved with time.

As a consequence of trial averaging, the vector plot is much less susceptible to the high trial-by-trial variability of the neuro-

$P(d \mid r, \text{cond}_{\text{unknown}}) \approx \frac{P(d) \cdot P(r \mid d, \text{cond}_{\text{unknown}})}{P(r \mid \text{cond}_{\text{unknown}})}$, \hspace{1cm} (13)

where all available training data are used to approximate the probability of a direction when the context is unknown, for the context unknown case, and

$P(d \mid r, \text{cond}_{\text{d}}) \approx \frac{P(d) \cdot P(r \mid d, \text{cond}_{\text{d}})}{P(r \mid \text{cond}_{\text{d}})}$, \hspace{1cm} (14)

where training data from another condition are used to approximate the probability of a direction with a different context ($d'$), for the generalization case.

The inaccuracy $T_A$ of time inference was quantified as the absolute difference between estimated task time $T$ and the true time $t_{\text{true}}$ from which test data $r_i$ were taken. Average accuracy in task time inference at each $t_{\text{true}}$ was computed across all single trials by leave-one-out cross-validation and then taking the mean across all test trials. Chance level was computed from the average inaccuracy achieved with random values assigned to $T$ (chosen randomly from the 1600 ms time span from 300 ms before PS until 300 ms after RS).

Results

Neuronal tuning is a function of time

In each of two monkeys, we analyzed the activity of 112 and 110 neurons, respectively, recorded under three different experimental conditions (Fig. 1). Typically, single neurons exhibited a marked temporal modulation of their spiking activity after the PS, which provided more or less prior information about the final movement target presented with the RS. Figure 2 shows a typical example of single-neuron activity under these conditions. As for most neurons, the trial-averaged rate profile (Fig. 2B, black curves) showed a clear dependence on movement direction, both during movement execution (interval from RS to RS plus 300 ms) and during the preparatory period (1 s from PS to RS) preceding the movement. Note, however, that the single trial spike responses (Fig. 2A) and rate profiles (Fig. 2B, gray curves) were highly variable across trials.

The time-dependent directional tuning of the mean firing rate of a neuron was visualized with the tuning vector (i.e., the vectorial sum of the trial-averaged firing rates from each movement direction executed after RS) (see Materials and Methods). During the time course of the trial, the resulting vector plot expresses the temporal evolution of both, the amplitude of tuning (vector length), and the preferred direction (vector direction) (for examples, see Fig. 3A, top; B, C, top panels). Before PS, the monkey had no information about the movement target, and thus, neuronal activity could not have been tuned to target direction. This resulted in very small residual amplitudes and random directions of the tuning vector, reflecting the estimation bias attributable to limited sample size (Fig. 3A–C, top panels). After PS, the vector amplitude increased and the vector direction stabilized as the tuning of the neuron evolved with time.

As a consequence of trial averaging, the vector plot is much less susceptible to the high trial-by-trial variability of the neuro-
nal activity. To account for the variability across trials in a statistical analysis of tuning based on single trials, we compared three different measures (see Materials and Methods), each of which provides us with a time-dependent estimate of the tuning strength of the neuron, which we refer to as the tuning profile of the neuron.

The first measure assesses the reliability of the tuning vector by computing at each point in time its \( z \)-score \( z_A(t) \) as the quotient of the mean amplitude and its SD \( \sigma_A(t) \) across 50 random single-trial sextets; (3) \( z \)-score measures tuning strength as the ratio of the vector length, \( |A| \), and \( \sigma_A \); (4) angular deviation \( s_A \) measures uncertainty in preferred direction across 50 single-trial sextets in degrees; the dotted line represents the chance level (\( s_A = 81\% \)); (5) decoding probability \( P_c \) for an arbitrary single-trial (solid line) and trial-averaged firing rate profile (dash-dotted line); the dotted line again represents the chance level (\( P_c = (1 - 1/6) \)).

The first measure assesses the reliability of the tuning vector by computing at each point in time its \( z \)-score \( z_A(t) \) as the quotient of the mean amplitude and its SD \( \sigma_A(t) \) across 50 random sextets (one for each of six directions) (see Materials and Methods) of single trial rate profiles. Values >1 indicate strong and reliable tuning (mean amplitude larger than trial-by-trial SD). The example neuron in Figure 3A shows values clearly larger than 1 during most of the preparatory period (\( z_A^{max} = 3.0 \)) and again during movement execution (\( z_A^{max} = 1.6 \)). The average \( z \) scores for all neurons in all three conditions are given in Table 1.

Note that our results for all tuning measures during the preparatory period remained qualitatively the same when computed with respect to the stimulus presented with PS instead of the movement executed after RS (see supplemental material, available at www.jneurosci.org).

### Table 1. Different tuning measures averaged across all neurons during the preparatory period from PS + 150 ms to RS − 150 ms, and during the execution period from RS to RS + 300 ms

<table>
<thead>
<tr>
<th>Tuning measure</th>
<th>PS + 150 ms to RS − 150 ms</th>
<th>RS to RS + 300 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One target</td>
<td>Two targets</td>
</tr>
<tr>
<td>( z_A )</td>
<td>1.13; 0.92</td>
<td>0.88; 0.76</td>
</tr>
<tr>
<td>( s_A )</td>
<td>53; 57</td>
<td>58; 61</td>
</tr>
<tr>
<td>( P_c )</td>
<td>0.199; 0.188</td>
<td>0.188; 0.182</td>
</tr>
</tbody>
</table>

Results are given separately for each behavioral condition. The first number in each entry is for monkey 1, and the second number is for monkey 2.
The second measure assesses the angular deviation \( s_A(t) \) of the tuning vector computed across the preferred directions of 50 sample vectors, constructed once more from randomly chosen single-trial sextets. This measure translates the trial-by-trial variability of the firing rate into an angular uncertainty of the directional tuning. The neuron selected for Figure 3A, analyzed in the complete information condition, exhibited an angular deviation \(<30^\circ\) during most of the preparatory period. This deviation was relatively small compared with the 60° angle between two adjacent movement targets, signaling strong tuning. The average angular deviations are provided in Table 1.

As a third measure, we computed the decoding probability \( P_c \) (i.e., the Bayesian probability to correctly decode the direction of the movement executed after RS from the single-trial firing rates) (see Materials and Methods). Our decoding model was repetitively trained with all trials from one out of the three behavioral conditions, except for one test trial (cross-validation) (see Materials and Methods). In Figure 3A, the decoding probability averaged across all test trials \( P_c(t) \) rose shortly after PS from \( P_c \approx 0.2 \) to a broad maximum of \( P_c \approx 0.33 \) in the middle of the preparatory period. It dropped to values \(<0.2\) around RS, after which it rose again to peak \( P_c \approx 0.23 \) during the actual movement. The average decoding scores \( P_c \) are given in Table 1.

Comparing panels 3–5 in Figure 3A, one observes that all three measures of tuning, \( z \) score \( z_A \), angular deviation \( s_A \), and decoding probability \( P_c \), measured the tuning profile of this neuron in a consistent manner: during the control period before PS, the absence of information about the forthcoming movement target was correctly reflected by a \( z \) score close to zero and by an angular deviation and decoding probability at chance level. During both preparatory period and execution period, the three measures were strongly correlated, as indicated in Table 2. In the following analyses, we focused on \( P_c \) to study in detail how tuning strength and single trial decoding of movement direction depended on time and prior information.

Analysis of the decoding probability \( P_c(t) \) in all single neurons recorded from two monkeys showed that the profile of each neuron was unique and highly dynamic in time (for five examples, see Fig. 3A,B,C, top panel). The broad range of different tuning profiles found in our data is exemplified in the three examples shown in Figure 3B: a neuron that exhibited tuning only during the preparatory period (top), a neuron that exhibited tuning only during the execution period (middle), and a neuron that exhibited tuning during both phases (bottom).

**Neuronal tuning depends on prior information**

Thus far, we considered only one behavioral condition in which the final target was already presented by the PS. In this case, complete information on target direction was available to the monkey throughout the preparatory period. In the alternative cases, either two or three adjacent targets were presented at PS, of which only one was chosen randomly to be presented as RS (Fig. 1). In these cases, therefore, during the preparatory period prior information about final target direction was incomplete.

We found that the tuning profiles of neurons were not only highly dynamic with respect to time, but also depended strongly on the amount of prior information: For example, the neuron depicted in Figure 3C exhibited clear tuning during the preparatory period, but no tuning during movement execution in the one-target condition (top panel). In both conditions of incomplete previous target information (bottom two panels), however, the neuron exhibited tuning after RS: there, the decoding probability \( P_c \) rose to 0.23 around movement onset (MO) in the two-target condition and to 0.27 in the three-target condition. During the preparatory period, however, this relationship was reversed: \( P_c \) decreased from a broad maximum of 0.33 in the middle of the preparatory period in the one-target condition to somewhat lower maxima of 0.29 and 0.23 early in the preparatory period in the two- and three-target conditions, respectively. This behavior suggests that this neuron dynamically allocated its processing power according to the availability of directional information.

Such dynamic behavior was also reflected in the population average of \( P_c \) across all 112 (monkey 1) and 110 (monkey 2) single neurons. Figure 4A shows that, starting 150 ms after PS and throughout the remaining preparatory period, the average decoding probability reflected the amount of prior information: \( P_c \) was highest in the complete information condition (black curve)
and lowest in the three-target condition (light gray). After RS, the condition of complete prior information differed from the conditions of incomplete prior information. For complete prior information, the maximum in $P_c$ was shallower for both monkeys and was reached earlier in monkey 1. For incomplete prior information, the rise in $P_c$ was steep and reached a higher maximum value in both monkeys ($p < 0.001$, Wilcoxon’s rank sum test).

Again, the alternative measures of tuning strength, $z$ score, and angular deviation, confirmed this behavior in both monkeys (Table 1).

Previous findings suggested a net flow of movement relevant information from premotor areas to M1 (Shen and Alexander, 1997b; Kakei et al., 2001, 2003; Cisek, 2006). Consistent with this hypothesis, we found that the single-trial decoding probability $P_c$ in Figure 4A reached its initial plateau earlier (150, 160, 180 ms after PS for one, two, and three targets) in PMd (monkey 2) than in M1 (250 ms, all conditions, monkey 1). This is also reflected in an earlier increase of the average firing rates in monkey 2.

**Relationship between neuronal tuning and firing rate**

We found that the neuronal tuning and the average firing rates across directions followed quite different behaviors in time. First, the time-resolved decoding probability $P_c(t)$ (Fig. 4A) and the time-resolved average firing rates (Fig. 4B) exhibited different temporal profiles: the rate peaked earlier, ~125–160 ms after PS, and then dropped to the same level in all three information conditions. In contrast, $P_c$ peaked ~150–250 ms after PS and then stayed at a different level for each of the three prior information conditions. Also, in monkey 1, in the complete information condition, $P_c$ rose to a second peak around RS, whereas the rate remained constant.

Second, the temporal dynamics of rate profiles and decoding probabilities showed remarkable variations across individual neurons and across prior information conditions (see examples in Fig. 3). The neuron in the bottom panel of Figure 3B, for example, strongly increased its tuning during movement preparation, whereas its firing rate averaged across directions remained relatively constant. The neuron depicted in Figure 3C behaved differently: During the preparatory period of the complete information condition, firing rate and decoding probability increased and decreased in parallel, whereas in the two- and three-target conditions tuning became weaker or even dropped to chance level, whereas the rate remained as in the complete information condition. During the movement period of the two- and three-target conditions, the relationship of rate and tuning differed again: now, tuning rose and peaked, whereas the rate decreased or remained at a low level.

Third, the occurrence of high decoding probability was only weakly correlated with high firing rates. In supplemental Figure 1 (available at www.jneurosci.org as supplemental material), we show scatter diagrams of the relationship between firing rate and $P_c$ at two points in time (150 ms after PS and 150 ms after RS). High values for $P_c$ occurred for a broad range of firing rates. Vice versa, high average firing rates resulted in a wide range of decoding probabilities. Across all recorded neurons, firing rates and $P_c$ were weakly, but significantly, correlated with values between 0.15 and 0.45, depending on time, condition, and monkey (supplemental Fig. 1, available at www.jneurosci.org as supplemental material).

There are several possible explanations for these observations. For example, an increase in decoding probability could result from an increase of the firing rate for the preferred direction and a decreased rate for the antipreferred direction, leading to no change in the average firing rate. Or a more dispersed distribution of firing rates across trials (i.e., a larger trial-by-trial variability) would lead to more uncertainty of the previous probabilities and thus to a reduced decoding probability even for a fixed average firing rate. Likewise, less variability along with reduced firing rates could lead to increased decoding probabilities with decreasing average rates (see Discussion).

**Distribution of tuning strength and tuning duration**

To quantify the contribution of individual neurons to the tuning properties of the neuronal population, we examined the distribution of tuning strengths and tuning durations in the recorded neural populations. The time-resolved distribution of decoding probabilities $P_c$ shown in Figure 5 reveals that at any point in time after PS only a fraction of neurons exhibited values of $P_c$ above chance level. The percentage of neurons exhibiting a significant $P_c$ ($p < 0.001$) (see Materials and Methods) at any given point in time is indicated by the black curves in Figure 5. During the preparatory period, this fraction increased monotonically with the amount of prior information: one-target condition, 50% (top panel); two-target condition, 40% (middle panel); three-target condition, 20% (bottom panel). For monkey 2, this fraction rose from ~25, 20, and 15% (one-, two- and three-target conditions, respectively) at the start of the preparatory period to approximate levels of 50, 40, and 30% at its end. After RS (i.e., during move-
ment execution), we found for both monkeys that the fraction of neurons with a significant decoding probability was higher in the two- and three-target conditions than in the condition of complete prior information. At the same time, also the absolute values of $P_c$ were higher for two or three presented targets than for one target only.

We saw that individual neurons exhibited a characteristic temporal modulation of their tuning strength (see temporal profiles in Fig. 3). Therefore, the contribution of each individual neuron to the population average (Figs. 4, 5) changed over time. To quantify the single neuron contribution statistically, we counted for each cell the uninterrupted episodes of significant tuning ($p < 0.001$, see Materials and Methods) and measured their durations. For this analysis, we pooled all neurons from both monkeys. We found that, in any of the three information conditions, $<10\%$ of all neurons never reached significance throughout the task. Generally, the numbers of significant episodes per neuron (Fig. 6A) were similar for the three behavioral conditions: the majority of the neurons exhibited one or two episodes of significant tuning (55, 56, and 53\% of all neurons in the one-, two-, and three-target condition, respectively). Their duration of uninterrupted episodes varied with the behavioral condition (Fig. 6B): most often it was quite short (mean duration $<300$ ms in 63, 73, or 86\% of all neurons in the one-, two-, or three-target condition, respectively). For complete target information, a considerable fraction of 12\% of all cells exhibited long tuning episodes, lasting for at least 900 ms. This means that these neurons stayed tuned during a long fraction of the preparatory period, and also during most of the movement execution phase. One such example is shown in Figure 3B (bottom panel). In the two- and three-target conditions, in contrast, in which movement preparation could only be completed after RS, the fraction of neurons exhibiting long tuning episodes ($>900$ ms) dropped to 7 and 2\%, respectively. Together, the distribution of tuning strength and tuning duration in the recorded neurons shows that (1) at any time only a few neurons are significantly tuned, and (2) individual neurons are tuned at different instances in time. Thus, directional information is distributed across a subset of neurons and the composition of this neuronal ensemble varies over time.

Effect of prior information on population coding
How accurately can we infer the direction of movement during movement preparation and execution from single-trial activity of a population of neurons? To measure this, we constructed a pseudopopulation by randomly selecting $N$ neurons from all recorded single neurons in one monkey. For a given degree of prior information, we then randomly selected one trial per neuron, each corresponding to the same movement direction. From this $N$-tuple of single-trial rate profiles, we estimated the average time-resolved decoding probability $P_c(t)$. This procedure was then repeated with different nonoverlapping subsets of $N$ neurons to obtain the average decoding probability for a given population size $N$ (see Materials and Methods). By varying the number $N$, we measured the dependence of decoding probability on population size; results are shown in Figure 7.

As expected, $P_c$ increased monotonically with neuron population size (Fig. 7A). During the preparatory period, $P_c$ is naturally bounded by the amount of prior information available to the monkey. Thus, in the three-target case, the final direction was predetermined with a probability of 1/3 (and 1/2 in the two-target case, respectively) as indicated by the dash-dotted lines in Figure 7. In all three prior information conditions, a population size of $\sim 50–100$ randomly selected neurons was sufficient to reach this upper bound already shortly ($<200$ ms) after PS, the directional information remained at this level throughout the preparatory period. The average $P_c$ during the preparatory period (beginning 150 ms after PS) for 100 neurons in the complete information condition was $0.90$ ($0.79$) for monkey 1 (monkey 2). In the two- and three-target conditions, decoding probabilities of 0.49 ($0.42$) and 0.30 ($0.29$) were achieved. Shortly after the RS, when complete target information was available in all conditions, the decoding probabilities assumed values $>0.98$ for both monkeys in all conditions. Note the increased variations in the decoding probability for larger population sizes. This is because these curves were necessarily based on averages over fewer pseudopopulations, resulting in noisier curves. Note also that, in our analysis without error trials, $P_c$ values near the upper bound in all three conditions (Fig. 7A) also indicate that the motor cortex accurately represented the instruction stimuli of all three conditions. $P_c$ values for stimulus decoding here reached values close to unity also for the two- and three-target conditions (supplemental Fig. 3, available at www.jneurosci.org as supplemental material).

Taking account of the distribution of tuning strength and tuning duration (Figs. 5, 6), the decoding probability of the neuronal population is based on neurons of which only a fraction of $\sim 20–50\%$ is significantly tuned at any point in time. Of these neurons, only a few showed a high level of tuning strength, whereas most neurons were weak encoders of movement direction [consistent with previous results (Wessberg et al., 2000; Mehring et al., 2003; Hatsopoulos et al., 2004)]. To illustrate the potential gain by selecting a small neural ensemble, we picked the 10 neurons that exhibited the highest time-averaged $P_c$ during the preparatory period and found that their decoding probability was comparable with that of 50 randomly chosen neurons (supplemental Fig. 4, available at www.jneurosci.org as supplemental material).
For all decoding probabilities computed thus far, the trials comprising the training data were taken from the same prior information condition as the test trials (i.e., the context was assumed to be known). With regard to a more general approach, such decoding strategy requires that the condition of the test trial is known and training data for this condition are available. In practical applications (e.g., for brain–machine interfacing), this requirement may not always be satisfied. We therefore examined the decoding probability also for two alternative scenarios.

First, we investigated generalization. For this, we computed the decoding probability when the training trials were exclusively taken from a different experimental condition than the test trials. This reflects the case in which a decoding algorithm is forced to generalize from one particular experimental condition (training data) to a different condition (test data). Figure 7B shows that generalization (red curves) led to a considerably decreased decoding probability compared with the case of a known context (black curves), both during preparation and execution of movement.

Second, we computed the decoding probability when the training trials were pooled from all three prior information conditions (context unknown). This reflects the case in which the decoding algorithm was trained with data from multiple conditions, but the specific context of the test trial is unknown. We found that the decoding probability (Fig. 7C, green curve) was again lower (but less so than compared with the generalization scenario) (Fig. 7B) than in the case of a known context (black) in which training and test trials were taken from the same condition only. Thus, the threefold increase in the number of training trials obtained by pooling over conditions could not fully compensate for the increased variability in the training set. With a population size of 50 neurons, however, decoding probabilities were nearly as good as in the control case. Only for test trials belonging to the complete information condition, the decoding probability was still considerably reduced during the preparatory phase.

Note that, for the analyses in Figure 7, B and C, trials were aligned to MO rather than to the PS, because we wanted to compare $P_c(t)$ during the movement period, and to better eliminate the latency variability across trials and information conditions (see Materials and Methods). For the case of a known context (black curves), the alignment to either MO (Fig. 7B, C) or PS (Fig. 7A) yielded approximately equal decoding probabilities.

The difference in performance of our decoding procedure for different training scenarios is summarized in Figure 8. It shows that both scenarios, generalization and, to a lesser degree, context unknown, consistently lead to decreased decoding probabilities compared with the standard context known scenario. However, for higher numbers of neurons, the performance for context unknown was almost as good as for context known, and much better than for generalization.

Decoding task time and movement direction

Our data show that both single neuron activity and tuning strength typically change in time during a trial (see sample neu-

![Figure 7](https://example.com)
the strong time-dependent characteristics of neuronal tuning properties.

Discussion

Dynamic encoding of movement direction

In both monkeys, single neuron tuning profiles could not easily be sorted into clear-cut categories with respect to number, starting time, and duration of significantly tuned episodes but rather described a continuum. Such heterogeneity of tuning properties has been previously described in the premotor and primary motor cortex (Shen and Alexander, 1997a,b; Cisek and Kalaska, 2005; Churchland and Shenoy, 2007). Here, we showed that this heterogeneity extends into the time domain. Statistically, most neurons exhibited one or two rather short (≥250 ms) uninterrupted episodes during which the preferred direction typically remained stable. In both monkeys, a fraction of ~10% of all neurons showed an early increase in tuning strength shortly after PS, and then stayed tuned above chance level throughout the trial. This behavior is reminiscent of neurons involved in working memory, as described, for example, in prefrontal cortex (for review, see Romo and Salinas, 2003). In our experiments, however, the monkeys did not have to actively memorize the visual target cue, as it was present throughout the preparatory period. The tuning profiles of individual neurons were strongly modulated by the prior information about the upcoming movement. This was true also during movement execution (Fig. 3C), even though the RS and the executed movement were the same in all three conditions. This suggests that movement representation in the motor cortex is not organized according to “one movement—one code” and extends prior results on the influence of contextual information on movement parameter encoding (Georgopoulos et al., 1989; Riehle et al., 1997; Bastian et al., 1998, 2003; Hepp-Reymond et al., 1999; Roux et al., 2003; Cisek and Kalaska, 2005).

Combining both aspects of directional tuning, its modulation in time, and its dependence on prior information, leads us to the following interpretation of our data: During a delayed reaching task, computational demands vary with time. When the initial information is made available with PS, movement preparation starts and, in the case of complete target information, may be completed before RS. The initial peak of the average firing rate observed 150–200 ms after PS (Fig. 4) may, thus, reflect a high computational load on cells involved in movement preparation (“preprocessing” neurons) (Riehle and Requin, 1989; Riehle et al., 1997; Bastian et al., 1998, 2003; Hepp-Reymond et al., 1999; Roux et al., 2003; Cisek and Kalaska, 2005).
www.jneurosci.org as supplemental material) and reached average probabilities $P_1 > 0.98$ for the correct decoding of movement direction around movement onset for 100 neurons. Throughout the preparatory period, the neuronal population accurately reflected the stimulus (supplemental Fig. 3, available at www.jneurosci.org as supplemental material) and, likewise, the degree of prior information about the final movement target (Fig. 7).

From a computational perspective, our results suggest that, after PS, the motor cortex prepares a movement as early and accurately as possible, based on the available information.

Cisek and Kalaska (2005) raised the question whether, in case of ambiguous target information, the motor cortex prepares alternative movements simultaneously, or whether it adopts a “guess and switch” rule, in which only one movement is prepared (guess), which after final target specification is possibly switched. In their experiment, the monkey was confronted with two opposite (i.e., conflicting) movement targets, and their analysis supports the former model of simultaneous encoding. In our experiment, two or three adjacent targets increased the angle of possible movement direction. Thus, target information was not causing a conflict, but rather an inaccuracy. This opens the possibility for an alternative strategy, in which the motor cortex prepares the movement less precisely in the case of incomplete information.

**Trial-by-trial variability and “optimal subspace hypothesis”**

Trial-by-trial variability of single-neuron activity in the motor cortex changes with time and in relation to behavior (Oram et al., 2001; Nawrot et al., 2003, 2008; Churchland et al., 2006; Nawrot, 2009; Ponce-Alvarez et al., 2009). Previous analyses of our data showed that, in both monkeys, the initially high single-neuron variability dropped sharply shortly after PS, and again shortly after RS (Nawrot et al., 2000, 2001; Rotter et al., 2005). Therefore, the increase in $P_1$ observed after PS and again after RS (Figs. 4, 5, 7) results from a combination of decreased variability and increased directional modulation of firing rates.

Churchland et al. (2006) formulated the optimal subspace hypothesis for movement preparation in the premotor and primary motor cortex. According to this hypothesis, each movement plan is represented by the activity of an adequate neuronal ensemble that occupies a subspace in the high-dimensional space of all possible activity patterns. After a target comes up, movement preparation leads to a gradually refined representation of the planned movement until the pattern of neuronal ensemble activity occupies the relevant subspace, resulting in the desired accuracy for a correct movement execution. The authors based their hypothesis on observations of neuronal variability matching with our results: variability decreased with the occurrence of a preparatory cue, and again with the GO cue. Additionally, we found in both monkeys an increase in variability during movement preparation from one target to two and three targets (supplemental Fig. 5, available at www.jneurosci.org as supplemental material) (Nawrot et al., 2009). Thus, our results support the optimal subspace hypothesis: in the two- and three-target condition, movement preparation could only be refined to a certain stage and, thus, occupied a subspace that is more extended than the optimal subspace associated with a unique movement direction.

**Interpretational issues**

There is strong evidence that processes associated with sensory-motor transformations are reflected in the activity of motor cortical neurons (Murata et al., 1997; Shen and Alexander, 1997a,b; Zhang et al., 1997; Kakei et al., 2003; Paz et al., 2003; Schwartz et al., 2004; Cisek and Kalaska, 2005). For example, it has been shown that, early in the preparatory period, neurons in PMd and M1 are preferentially tuned to the location of a visual target, and then gradually shift their tuning to movement direction (Shen and Alexander, 1997a,b; Zhang et al., 1997). Thus, the observed time-dependent involvement of neurons as discussed above may...
to some extent reflect a transformation from a stimulus representation in form of a visual target to a representation of motor output. It cannot be completely ruled out, however, that the recorded activities were, at least partially, influenced by additional factors, in particular the visual stimulus itself, eye movements, or attentional effects. However, we are convinced that our findings cannot be explained by these effects alone. The encoding of a visual stimulus could not explain why firing rates and tuning strengths are elevated at the end of the preparatory period, and why this late preparatory activity is correlated with reaction time (Riehle and Requin, 1993). Eye movements were reported to affect motor cortical activity (Boussaoud, 1995; Baker et al., 1999), but are unlikely to exert a major effect in our data since eye movements after PS were phasic and nonsystematic during the preparatory period (A. Riehle, unpublished observations). Similarly, attention is known to affect activity in premotor cortex (Boussaoud, 2001), but a dominant role of attention is contradicted by our finding of higher activities and less variability during the preparatory period in the complete information condition compared with both incomplete information conditions, in which attention would be expected to be higher.

Relevance for brain–machine interfaces

Our results suggest that, during movement preparation and execution, single-trial firing rates of 50–100 randomly chosen, or of ~10 selected (supplemental Fig. 4, available at www.jneurosci.org as supplemental material) motor cortical neurons could allow an essentially error-free decoding of a two-dimensional movement direction. These numbers lie within the previously reported range of ~10–20 neurons (Serruya et al., 2002; Taylor et al., 2002), and several tens to hundreds of neurons (Wessberg et al., 2000; Carmena et al., 2003; Musallam et al., 2004). In addition, we could show that it is feasible to jointly decode direction and task time at a significant cost in performance for direction decoding [complementing the results of Achtman et al. (2007)]. Both results could be of practical relevance for future BMI applications in patients (Hochberg et al., 2006). Under realistic conditions, the time until the execution of an intended movement could be an important additional parameter to be estimated. Also, it will likely be advantageous for subjects that operate a brain-controlled prosthesis to make full use of contextual information, because it may help planning and executing a desired (prosthetic) limb movement. A first step to improve decoder performance could be its training in different context settings. In our off-line analyses, training with the pooled data from all three information conditions (context unknown) (Fig. 7B) allowed to partly compensate for the lack of knowledge about the actual condition. Models for decoding complex movements therefore might have to take into account various contextual parameters to successfully operate a BMI. Here, neuronal activity recorded from other brain areas may contribute to successfully decode context-related information (Shenoy et al., 2003; Musallam et al., 2004).

References


Supplemental Material

Correlation of firing rate and decoding probability

Supplement figure 1 gives the results for the estimation of the linear correlation between mean firing rates and decoding probability $P_C$ at two instances in time: (1) in the early period of the preparatory phase 150ms after the preparatory stimulus (PS) was presented, and (2) 150ms after the reaction signal (RS) was presented. Both time points correspond to the maxima of the average firing rates (cf. Fig. 4).

**Supplemental Figure 1:** Correlation of firing rate and decoding probability. Each panel shows the average firing rate scattered against the decoding probability of individual neurons, both plotted on log-scale. Each dot represents one neuron. The distribution of dots depicts rates and decoding probabilities for all neurons of monkey 1 (top) and monkey 2 (bottom), either during movement preparation (left) or movement execution (right). Different symbols represent neurons recorded during the complete information condition (+), the 2-target (x), and the 3-target condition (+).
Stimulus coding

In the main text, throughout the whole trial, we decoded the direction of the executed movement. Here, we take a different perspective and, during the preparatory period, decode the preparatory stimulus PS (see Materials and Methods for details). As we considered only correctly executed trials, this analysis made no difference for the 1-target condition. In the 2-, and 3-target conditions we have 3 or 2 preparatory stimuli (PS) and thus the decoding reduces to a 3- or 2-class problem, respectively.

First, we repeated our analyses of single neuron tuning strength. The results are summarized in supplementary table 1. One can see, that tuning strengths as computed by $Z_A$ and $S_A$ decrease from 1-target to 2-target to 3-target – as they do when tuning strength was computed with respect to the executed movement. $P_c$ is difficult to compare across conditions here, because its baseline increases from $1/6$ to $1/3$ to $1/2$.

<table>
<thead>
<tr>
<th>Measure</th>
<th>1-target</th>
<th>2-target</th>
<th>3-target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_A$</td>
<td>1.13</td>
<td>0.76</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.92</td>
<td>0.64</td>
<td>0.51</td>
</tr>
<tr>
<td>$S_A$</td>
<td>53</td>
<td>61</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>57</td>
<td>63</td>
<td>64</td>
</tr>
<tr>
<td>$P_c$</td>
<td>0.199</td>
<td>0.377</td>
<td>0.532</td>
</tr>
<tr>
<td></td>
<td>0.188</td>
<td>0.364</td>
<td>0.530</td>
</tr>
</tbody>
</table>

**Supplementary table 1:** Different measures for tuning to the preparatory stimulus, averaged across all neurons during the preparatory period from PS+150ms to RS-150ms, Results given separately for each behavioral condition. First number in each entry is for monkey 1, second number is for monkey 2.

If we look at the number of significantly tuned neurons computed from those neurons with significant $P_c$ ($P<0.001$, see Materials & Methods), we find a decrease from 1-target to 2-target to 3-target though (cf. supplementary figure 2) – as we found on the basis of $P_c$ for decoding the executed movement (Figure 5). For monkey 1, the average number of significantly tuned neurons from PS+150ms until RS drops from 50% in the 1-target condition to 37% in the 2-target condition to 18% in the 3-target condition. For monkey 2 this number drops from 30% to 27% to 16% likewise.

Supplementary figure 2 also shows that the distribution of $P_c$ values looks similar to their distribution when $P_c$ was computed for decoding the executed movement shown in Figure 5. At any point in time, most neurons, indicated by red and yellow colors, are not or only weak encoders of the stimulus. Few neurons, indicated by blue colors, at any point in time are strong encoders of the stimulus.
Supplemental Figure 2: Time-resolved distribution and proportion of significant P_c values for decoding the preparatory stimulus (population of 112 neurons in monkey 1). Otherwise as in Figure 5 of the main text.

Second, we decoded the PS during the preparatory period from a population of 100 neurons. The Bayesian probability P_c for decoding the correct stimulus from a population of 100 neurons in monkey 1 reaches values close to unity (>0.85) shortly after PS in all three conditions (supplementary figure 3). Monkey 2 yields somewhat weaker values (>0.6, not shown). As to be expected, the stimulus type (1 target LED,
2 target LEDs, or 3 target LEDs) does not influence the quality of stimulus decoding. These results are consistent with and could be expected from our results on the population decoding of movement direction in Fig. 7. There, we also found decoding probabilities for the final movement direction that were close to the upper bound with \( P = 1 \), \( P = 1/2 \) and \( P = 1/3 \) for 1, 2, and 3 targets presented with PS (see also Materials and Methods). Together with the result in Fig. 7 we can conclude that a population of 50-100 neurons carries information about the stimulus/direction of movement that comes close to the maximal information.

Supplemental Figure 3: Population decoding of stimulus. Bayesian probabilities for decoding the correct preparatory stimulus (PS) from a population of 100 neurons. In the single target condition there were 6 possible instruction stimuli with chance level 1/6. In the 2- and 3-target condition there were 3 and 2 possible stimuli, respectively (see Materials and Methods). Decoding statistics as in Fig. 7.

Note, that with the additional analyses on stimulus coding presented here, we do not intend to suggest that cells in M1 or PMd encode visual stimuli per se. The presentation of the visual stimuli alone without the association of any behavioral relevance should lead to no or diminished motor cortical responses.

Population decoding from a selected set of neurons

The distributions in Fig. 5 or suppl. Fig. 2 show that at any point in time only a subset of neurons are significantly tuned and of those only few neurons show strong tuning. Moreover, individual neurons are significantly tuned mostly during short periods of time (Fig.6). Consequently, at any point in time a different subpopulation of all neurons contributes to the population decoding. Here, we selected 10 neurons that showed the highest time-averaged decoding probabilities and repeated the time-resolved decoding from their single-trial activities. As shown in the supplemental Fig. 4, this subset of 10 selected neurons reached decoding probabilities similar to a set of 50 randomly chosen neurons (cf. Fig. 7).
Supplemental Figure 4: Decoding probability for the 10 'average best' neurons (monkey 1). Computation performed as in Fig. 7, except that here the 10 neurons with highest time-averaged Pc as computed during preparatory and execution period were used. Results for complete information condition (black), 2-target condition (dark gray), and 3-target condition (light gray). Dotted line depicts chance level.
Dynamic changes of trial-by-trial variability

We performed a time-resolved analysis of trial-by-trial spike count variability. To this end we measured for each neuron the Fano factor, i.e. the count variance normalized by the mean count across trials, in a sliding window as described elsewhere (Nawrot et al., 2008; Nawrot, 2009). Supplemental Fig. 5 shows the average Fano factor across all neurons of monkey 1 in all three conditions. We found that before PS the variability was highest in the 3-target condition and lowest in the 1-target condition. Note, that the monkey performed the tasks associated with the 3 different conditions in separate blocks. The order of blocks was randomized across different recording sessions. During the preparatory period trial-by-trial variability reached the lowest value in the condition where prior target information was complete, and it was higher in the conditions with incomplete prior information.

**Supplemental Figure 5**: Temporal modulation of trial-by-trial variability depends on prior information. Time resolved Fano factor calculated in a moving window of 450ms width (indicated by black intervals) for the three different conditions of prior target information and averaged across all 112 neurons recorded in monkey 1. Error bars indicate standard error of the mean. Black: 1-target condition, red: 2-target condition; blue: 3-target condition.